



高等学校图书馆学核心课教材

数字图书馆原理与技术

主 编 黄如花
副主编 王 梅 黄晓斌



WUHAN UNIVERSITY PRESS

武汉大学出版社

高等学校图书馆学核心课教材

数字图书馆原理与技术

主 编	黄如花				
副主编	王 梅	黄晓斌			
编著者	黄如花	王 梅	黄晓斌	胡永生	
	陈 朋	韩 丽	冯艳花	武利红	
	李 瑜	向瑕嫫	李鹏云		

武汉大学出版社

图书在版编目(CIP)数据

数字图书馆原理与技术/黄如花主编;王梅,黄晓斌副主编. —武汉:武汉大学出版社,2005.9

高等学校图书馆学核心课教材

ISBN 7-307-04557-5

I. 数… II. ①黄… ②王… ③黄… III. 数字图书馆—基本知识 IV. G250.76

中国版本图书馆 CIP 数据核字(2005)第 067123 号

责任编辑:严红 金鑫 汪芳 康苗 责任校对:黄添生 版式设计:支笛

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:武汉市新华印务有限责任公司

开本:880×1230 1/32 印张:13.125 字数:360千字 插页:2

版次:2005年9月第1版 2005年9月第1次印刷

ISBN 7-307-04557-5/G·741 定价:21.00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。

内 容 提 要

本书是武汉大学“十五”规划教材，也是图书馆学核心课程系列教材之一。

本书系统地阐述了数字图书馆的相关概念、数字图书馆产生与发展的社会背景、国内外数字图书馆的研究与建设情况、数字图书馆的组织与实施、数字图书馆的构建、数字信息资源的建设与处理（包括元数据、RDF、XML、数字信息组织及其优化）、数字图书馆信息检索、数字图书馆信息服务（数字参考服务）、数字信息的长期保存与维护、数字图书馆的应用、数字图书馆的经济法律与社会问题、数字图书馆人力资源管理以及数字图书馆教育等问题。

本书内容较为全面、详实，书后附录实用性强，适合于作为图书馆学专业、信息管理与信息系统专业和其他相关专业的教材，也可供数字图书馆规划、建设与开发的有关人员参考。

本书受到了国家社会科学基金（项目代号03BTQ021）和武汉大学“十五”教材规划的资助。

前 言

20 世纪末以来,以信息技术为代表的新技术革命浪潮席卷全球,深刻地影响着世界经济和科技的发展格局,不断改变着人们的工作、学习、思维和生活方式。以海量信息的有序组织为特点的数字图书馆理念与技术,采用了下一代因特网上信息资源的管理模式,可以从根本上改变目前因特网上信息分散、使用不便的现状,因而被世界各国视为国家信息基础设施(NII)建设、知识创新体系和创新能力的重要组成部分。各国政府机构、各类图书馆、高等院校、科研机构及与之相关的企业纷纷斥资,对数字图书馆进行研究与开发,取得了丰硕的成果。

数字图书馆的建设需要一大批熟练掌握数字图书馆理论与实践技术的高级专门人才。我们通过网上调查发现,国内外近几年开设数字图书馆课程的高等院校不断增多,国外使用的英文版数字图书馆教材达 30 余种,而国内数字图书馆著作中适合做教材者为数不多,本书正是为了适应国内图书情报学教育改革和数字图书馆建设的需要而编写的本科生教材。与同类著作相比,本书的特色主要体现在:①内容全面。力求涵盖数字图书馆涉及的基本知识、基本理论、基本技能、数字图书馆建设过程中的重要环节和国内外重要的数字图书馆项目。②注重理论与实践的结合。在理论阐述的同时,注重穿插国内外典型数字图书馆案例

的介绍与分析,以增强其对实践的指导作用。③资料丰富。在本书的撰写过程中,我们广泛吸取了国内外有关数字图书馆理论与实践的研究成果(包括图书、期刊论文、会议论文、学位论文和大量的网上资源),每章末配有思考题,并把多年从事数字图书馆有关课程教学与科研中积累和经常参考的资料作为附录置于书后,以便与广大读者共享。

全书共十一章,主要内容包括:数字图书馆的概念、数字图书馆产生与发展的社会背景、国内外数字图书馆的研究与建设、数字图书馆的组织与实施、数字信息资源的建设与处理、数字图书馆信息检索、数字图书馆信息服务(数字参考服务)、数字信息的长期保存与维护、数字图书馆的应用、数字图书馆运行模式和知识产权保护等问题、数字图书馆人力资源管理、数字图书馆教育。本书由黄如花负责全书的结构设计、各章内容修改和全书的最后定稿。武汉大学图书馆王梅研究馆员和中山大学信息管理系副教授黄晓斌博士担任副主编,王梅研究馆员还参与了全书的结构设计工作,武汉大学信息管理学院从事数字图书馆和信息检索与服务方向的多位研究生参与编写。具体分工情况如下:第一章由冯艳花执笔,第二章由武利红执笔,第三章由胡永生执笔,第四、五章的第一、二节由王梅执笔,第四章的第三节和第五章的第三节由黄晓斌执笔,第六章的第一、二节由李鹏云执笔,第三节由黄如花执笔,第七章由陈朋执笔,第八章由黄如花执笔,第九章由韩丽执笔,第十章的第一、二、四、五节由李瑜执笔,第三节由向瑕嫔执笔,第十一章第一节由胡永生执笔,第二至五节由黄如花执笔,书后参考文献和附录1~3由黄如花执笔。研究生吴刚同学参与了附录1中部分外文资料的编译工作。修改时,根据需要对有关内容做了增删或调整,如第五章原“运行机制”部分的内容调整到第十章第一节,第五章第一节增加了第三部分,第十章中原关于数字图书馆的技术与标准化问题分别整合到其他有关章节,原第九章第一节的“信息服务”改为“学术信息交流”。限于篇幅,数字信息安全等内容被删除。

本书的出版得到了国家社会科学基金项目“网络信息组织模

式的优化研究”（项目代号 03BTQ021）和武汉大学“十五”教材规划的资助。感谢武汉大学信息管理学院领导和出版社社长陈庆辉教授的关心与大力支持。感谢中国数字图书馆工程建设联席会议办公室申晓娟副主任对附录 3 提出的宝贵意见。武汉大学出版社编辑室主任严红副编审为本书的出版付出了辛勤劳动，提出了许多宝贵的修改意见，在此表示诚挚的谢意！本书在写作过程中，参考了大量的文献（包含网上资源），限于篇幅，有的未能列出。在此，对所有这些参考文献的作者（列于书后的或未列出的）表示深深的敬意！

数字图书馆是一个发展很快的领域，本书难以及时反映国内外最新的进展，加之作者水平有限，对许多新内容的消化吸收做得不够，且多头执笔，虽几经统稿，书中不妥乃至错误之处在所难免，恳请专家、同行和读者批评指正。

编著者

2005 年 2 月于珞珈山

目 录

前 言	1
第一章 数字图书馆概论	1
第一节 数字图书馆的定义	1
第二节 数字图书馆的特征	7
第三节 数字图书馆的作用	11
第二章 数字图书馆建设与社会环境	17
第一节 技术的发展与进步	17
第二节 政府的支持和经济保障	24
第三节 用户信息素质的提高	27
第三章 数字图书馆的研究与建设	33
第一节 国外数字图书馆的研究与建设	33
第二节 国内数字图书馆的研究与建设	54
第四章 数字图书馆的组织与实施	72
第一节 数字图书馆组织模式	72
第二节 数字图书馆的实施	78
第三节 数字图书馆联盟	86
第五章 数字图书馆的构建	97

第一节	数字图书馆的体系结构	97
第二节	数字图书馆的功能设计	105
第三节	数字图书馆开发平台	109
第六章	数字图书馆信息资源的建设与处理	122
第一节	数字化信息资源的来源	122
第二节	数字信息资源的描述和处理	131
第三节	元数据与资源描述框架	145
第四节	数字图书馆信息组织的优化	166
第七章	数字图书馆的信息检索	168
第一节	数字图书馆信息检索系统的设计	168
第二节	数字图书馆信息检索方法	177
第三节	数字图书馆的集成化信息检索与联邦检索	186
第八章	数字图书馆的信息服务	199
第一节	开展数字图书馆信息服务的必要性	200
第二节	数字图书馆信息服务的内容	204
第三节	QuestionPoint 合作数字参考服务系统	216
第九章	数字图书馆的应用	223
第一节	学术交流	223
第二节	电子商务	225
第三节	远程教育	231
第四节	电子政务	236
第五节	电子出版	242
第十章	数字图书馆的经济、法律与社会问题	246
第一节	数字图书馆的经费与运行模式问题	246
第二节	数字图书馆建设与服务中的知识产权问题	258

第三节	数字信息的长期保存与维护问题	268
第四节	数字图书馆的人文与伦理问题	283
第五节	数字鸿沟问题	287
第十一章	数字图书馆人才培养	293
第一节	数字图书馆人才培养的模式	293
第二节	国内外在校教育中数字图书馆课程开设情况	297
第三节	国内外继续教育中数字图书馆课程设置情况	315
第四节	国内外短期培训中数字图书馆课程的设置	319
第五节	国内外数字图书馆课程开设情况的总结	324
附录 1	数字图书馆常用词汇中英文对照	333
附录 2	数字图书馆有关的网络资源	358
附录 3	国外数字图书馆研究专家	374
主要参考文献		391

图 表 目 次

表 1	国内外重要元数据简况	148
表 2	都柏林核心元数据历届会议情况	154
表 3	DC 核心元素及其分类	156
表 4	DC 限定词	158
表 5	数字图书馆信息检索系统用户接口子系统的功能	172
表 6	美国其他大学数字图书馆课程开设简况	298
表 7	祖国大陆图书馆学系（专业）本科阶段数字图书馆课程 开设情况（按学校名称字顺排列）	303
表 8	除美国之外其他国家开设数字图书馆课程简况	311
表 9	历届 Ticer 国际数字图书馆暑期研究班简况	320
表 10	历届美国数字图书馆联盟（DLF）论坛简况	321
表 11	使用较普遍的数字图书馆英文教材（按出版时间排 列）	327
图 1	数字图书馆生命周期模型	79
图 2	数字图书馆综合模型	79
图 3	数字图书馆建设的系统模型	80
图 4	数字图书馆体系的四个层面	103
图 5	数字图书馆系统组成要素	103
图 6	数字图书馆体系架构	104
图 7	数字图书馆的功能构架	106
图 8	RDF 的三元组结构	162
图 9	RDF 模型	164

数字图书馆原理与技术 | 图表目录

图 10	RDF 简单模型	164
图 11	信息检索系统的逻辑构成	170
图 12	Dienst 服务的交互模型	174
图 13	Inquirus2 体系结构	176
图 14	基于内容的视频检索流程	185
图 15	OAI 协议系统框架	189
图 16	OpenURL 链接流程图	191
图 17	MPEG-7 的一种描述表的结构	194
图 18	DNER 信息平台构建图	196
图 19	QuestionPoint 的合作模式图	216
图 20	QuestionPoint 的地区与全球系统构成	218
图 21	QuestionPoint 地区系统的工作流程	219
图 22	QuestionPoint 全球系统的工作流程	220

第一章

数字图书馆概论

第一节 数字图书馆的定义

一、数字图书馆定义的各种说法

1994年,美国国家科学基金会(NSF)、美国国防部高级研究计划署(DARPA)、美国国家宇航局(NASA)联合发起了“数字图书馆创始计划”(Digital Library Initiative,简称DLI),领导、组织和资助美国的数字图书馆研究与开发,把建立数字图书馆真正地作为一个研究领域。“数字图书馆”一词由此迅速被计算机科学界、图书馆界以及其他各领域所采纳。

对于什么是数字图书馆,目前国内外尚没有形成一致的看法。由于各自出发点不同,如有致力于基础理论和技术研究的,也有致力于应用方面研究的,加之数字图书馆又处于发展变化阶段,凡是参与了数字化信息资源研究与建设的群体和项目都各有自己的定义,包括图书馆界、数字图书馆项目建设者、计算机科学界、国际标准组织、国际组织(如联合国)等。

定义一:1995年召开的美国联邦信息基础设施技术与应用项目(Information Infrastructure Technology and Applications,简称IITA)数字图书馆专题讨论会所用的定义比较能代表当时官方和研究群体对数字图书馆的理解和期望。会议指出,“数字图书馆是向用户群体提

供便于查找利用庞大的、经过组织的信息和知识存储库的手段的系统。这个信息组织的特点是没有预知的关于信息使用的详情。用户进入这个存储库，重新组织和使用之。这种能力由于数字技术的能力而大大增强”。^①

定义二：在 1997 年 3 月由 NSF 赞助的关于“分散式知识工作环境”专题讨论会上，数字图书馆的概念得到扩充。会议报告指出，“‘数字图书馆’的概念并不仅仅是一个有着信息管理工具的数字收藏的等价词。数字图书馆更是一个环境，它将收藏、服务和人带到一起以支持数据、信息乃至知识的全部流程，包括从创造、传播、使用到保存的全过程”。^②

定义三：William Y. Arms 将数字图书馆非正式的定义界定为“有组织的信息馆藏及相关服务，信息以数字化形式保存，并通过网络进行访问”。^③

定义四：“数字图书馆一般而言是指利用当今先进的数字化技术，通过诸如 Internet 国际互联网等计算机网络，使人数众多且又处在不同地理位置的用户能够方便地利用大量的、分散在不同贮存处的电子物品的全部内容。这些电子物品包括网络化的文本、地图、图表、声频、视频、商品目录，以及科学、企业、政府的数据集，还包括超媒体和多媒体等。”^④

定义五：“一个数字化的信息系统，它将分散于不同载体、不同地理位置的信息资源以数字化方式贮存，以网络化方式互相连

① 曾蕾，张甲，杨宗英. 数字图书馆：路在何方？——关于数字图书馆定义、结构及实际项目的分析. 情报学报，2000（1）：66

② 曾蕾，张甲，杨宗英. 数字图书馆：路在何方？——关于数字图书馆定义、结构及实际项目的分析. 情报学报，2000（1）：66

③（美）William Y. Arms 著；施伯乐等译. 数字图书馆概论. 北京：电子工业出版社，2001：1

④ 杨向明. 21 世纪图书馆发展的方向——数字图书馆. 图书馆，1997（1）：45～46

接,提供即时利用,实现资源共享。”^①

定义六:“数字图书馆是一个大系统,它具有分布的、大规模的有组织的数据库和知识库,用户或用户团体可对系统内的数据库和知识库进行一致性的访问,获得自己所需要的最终信息。”^②

定义七:“数字图书馆是以数字形式存储和处理信息的图书馆,即对有高度价值的图像、文本、语言、音响、影像、影视、软件和科学数据等多媒体信息进行收集,组织规范性加工,进行高质量保存和管理,实施知识增值,并提供在广域网上高速横向跨库连接的电子存取服务,同时还包括知识产权、存取权限、数字安全管理范畴。因此准确地说,数字图书馆是超大规模的可以跨库检索的海量数字化信息资源库。”^③

定义八:“数字图书馆是以统一的标准和规范为基础,将有价值的文本、声音和图像等多媒体信息资源数字化,再实施知识信息增值加工后存储于分布式海量资源库群,并以智能检索技术为手段,以电子商务为管理方式,通过广域高速网络向全世界用户提供存储信息服务。”^④

定义九:“数字图书馆是社会信息基础结构中信息资源的基本组织形式,这一形式满足分布式面向对象的信息查询需要。”^⑤

定义十:美国数字图书馆联盟(Digital Library Federation)的定义广受认同,其认为:“数字图书馆是一种提供信息资源的组织,包括软硬件、网络与专业人员,以一致性、永久性的方式将数字化馆藏进行选择、组织、提供查询、解释、传播与完整保存,以便于这些数字

① 张丽虹. 数字图书馆及其相关问题与技术. 图书馆建设, 1997(2): 53

② 李人厚, 苗凌. 数字图书馆及其面临的挑战. 大学图书馆学报, 1997(6): 11

③ 徐文伯. 建设中国数字图书馆意义重大. 光明日报, 2000-3-8

④ 邱均平, 岳亚, 颜金莲. 论数字图书馆信息资源的知识产权管理. 资讯传播与图书馆学(台湾), 2000(3): 42

⑤ 刘炜, 刘年娣. 国内数字图书馆研究评述. <http://www.libnet.sh.cn/sztsg/fulltext/reports/1999/SZ1.html> (访问日期: 2005/1/5)

馆藏可以迅速、经济地提供给特定社区与人们使用。”^①

从以上定义不难看出，随着数字图书馆研究的不断深入，其涵义逐渐突破技术层面，涉及的范围越来越广，而信息的有组织性以及最大程度地满足用户需求是必不可少的。

二、数字图书馆与相关术语的辨析

(一) 数字图书馆与传统图书馆

“数字图书馆”一词是从 Digital Library (简称 DL) 翻译过来的，有的学者并不主张这种译法，因为数字图书馆中的“图书馆”不同于传统拥有大型书库、宽敞明亮的阅览室的实体图书馆，最准确的翻译应该是数字资料库^②。当然，数字图书馆与传统图书馆有着不可分割的联系，传统图书馆是数字化的源泉，目前走在数字图书馆研究和开发应用前列的仍旧是图书馆，国内外都是如此。我们可以从资源和服务模式两方面进行比较：

首先，传统图书馆以印刷文献为对象，重视物理形态的图书和期刊的收藏与保存，实行集中式管理，资料使用受地域限制。通过款目（以分类号或主题词为核心）对文献进行描述，相关书籍都按类排放在一起，读者阅读时以类索取。而数字图书馆收藏以数字化资料为主体，通过数字化减少物理形态资料的储藏和保存，实行分散式管理，使用不受地域限制；通过元数据对资源进行描述，内容比传统图书馆书目体系更加全面和复杂；存储突破空间的限定，相同主题的资料呈分布式状态，在物理意义上不必存放在一起。

其次，传统图书馆以印刷本资源为中心，提供固定模式的信息服务，用户获取信息是被动的，必须知道资料被哪个图书馆收藏，以及图书馆的位置与如何获取、利用等问题。数字图书馆以用户为中心，为具体用户提供个性化的信息服务。分布式存储技术的运

^① <http://www.diglib.org> (访问日期:2005/1/5)

^② 高文，刘峰，黄铁军等. 数字图书馆——原理与技术实现. 北京：清华大学出版社，2000：2

用,使用户不再受时间和空间的限制,无论在家、学校或办公室甚至在汽车里都能访问世界任何地方的数字图书馆。阅读基于超链接、关键词或任意经过定义的方式。

(二) 数字图书馆与图书馆自动化

图书馆自动化随着时代的不同而有不同的内涵。早期的图书馆自动化,旨在解决从出纳台到书库之间的借阅单和图书的传递问题,大大节省了图书管理人员往返于两地的时间。20世纪90年代,计算机技术和通信技术的发展使海量存储和高速传递成为可能,从此图书馆走向全面自动化的道路,无论是内部管理还是对外服务都实现了自动化。自动化集成系统将文献的采访、编目、流通等联成一体,书目查询、图书借阅、读者信息查询等均可通过计算机完成。

图书馆自动化强调图书馆业务管理的计算机化,而数字图书馆侧重于信息资源的数字化,突出信息的计算机存储处理、网络传输特点。从某种意义上说,图书馆自动化只是数字图书馆实现的前提,或者说数字图书馆是图书馆自动化的高级阶段。

(三) 数字图书馆与数据库

数据库是相互关联的数据集合,是知识与信息的数字化集合。数据库方式将某一主题的知识资源以固定的记录格式存储,并提供检索入口,用户通过检索入口,可以方便地找到所需的信息线索,进而通过信息线索直接链接到相应的网站。数据库还能高速处理大量结构化和非结构化的数据。如今的关系数据库增加了对图形、图像、声音、超文本等多媒体数据的存取、处理、管理功能,实现了从数据管理到对象管理的扩展,增加了知识管理的功能,面向对象数据库包含了更多的数据语义信息,对复杂数据对象的表达能力更强了。

数据库技术组织信息资源可极大地提高信息的有序性、完整性、可理解性和安全性。数据库技术与网络技术的融合极大地方便了用户利用和开发信息资源,提高了效率。Web数据库将数据库技术与Web技术很好地融合在一起,使数据库系统成为Web的重要组成部分,能够实现数据库与网络技术的无缝结合。