

二十一世紀圖書館學叢書 (第三輯)

任繼愈題

■ 丛书主编：丘东江

ZENYANG JIANSHE
SHUZI TUSHUGUAN

怎样建设 数字图书馆

刘晓清 主 编

宓永迪 陈 晔 刘柏嵩 张 杰 副主编



海洋出版社

怎样建设数字图书馆

	刘晓清	主编
宓永迪	陈 晔	
刘柏嵩	张 杰	副主编

海洋出版社

2010 年 · 北京

内容提要

本书以数字图书馆的应用为背景,以数字图书馆建设的实践为核心,比较系统地阐述了数字图书馆的信息环境、规划、系统构架、资源配置、版权保护、标准规范、系统安全、运营管理、绩效评估等重要部分。本书紧密联系当前数字图书馆建设和应用,根据数字图书馆发展的趋势,分析和概括了数字图书馆系统应用的各个环节,为数字图书馆的建设提供参考。

图书在版编目(CIP)数据

怎样建设数字图书馆 / 刘晓清主编. —北京:海洋出版社,2010.4
(21 世纪图书馆学丛书. 第三辑)

ISBN 978 - 7 - 5027 - 7646 - 6

I. ①怎… II. ①刘… III. ①数字图书馆 - 基本知识
IV. ①G250.76

中国版本图书馆 CIP 数据核字(2009)第 233036 号

责任编辑:杨海萍

责任印制:刘志恒

海洋出版社 出版发行

<http://www.oceanpress.com.cn>

北京市海淀区大慧寺路 8 号 邮编:100081

北京盛兰兄弟印刷装订有限公司印刷 新华书店发行所经销

2010 年 4 月第 1 版 2010 年 4 月第 1 次印刷

开本:850 mm × 1168 mm 1/32 印张:8.375

字数:225 千字 定价:30.00 元

发行部:62147016 邮购部:68038093 总编室:62114335

海洋版图书印、装错误可随时退换

《21 世纪图书馆学丛书》第三辑编委会

主 编：丘东江

编 委：王 波 刘晓清 祁延莉 姜火明

赵 志 高 红 黄飞燕 黄如花

赖雪梅

主 编 弁 言

《21 世纪图书馆学丛书》旨在顺应时代潮流,把握知识经济和网络时代图书馆学情报学发展脉络,强调精品意识和求实创新的学术风格,全力推出理论与实践相结合的最新研究著述,绝无新党八股文章。自第一、二辑出版以来,受到业内的好评和喜爱。

在编辑《21 世纪图书馆学丛书》第一辑之时,得到著名学者、时任国家图书馆馆长任继愈先生的关心和鼓励,并亲自为该丛书题写书名。

第三辑为 11 册,聘请具有丰富图书馆实际工作经验的学者专家以及能和实际工作相结合的教育工作者执笔,撰写选题实在、新颖、信息丰富、密切结合图书馆工作实践、值得广大图书馆工作者参考的专著。

《21 世纪图书馆学丛书》第三辑涵盖面较为广泛,所涉及范围包括《全球图书馆界的“联合国”——联机计算机图书馆中心(OCLC)》、《怎样建设数字图书馆》、《图书馆实用信息技术》、《图书馆信息共享空间建设》、《世界各国图书馆馆藏发展政策精要》、《图书馆员英语选读Ⅱ》、《图书馆学情报学经典著作选读》、《英文参考源的检索与利用》、《睁眼看世界——我们向国外图书馆学习什么》、《快乐的软图书馆学》以及《图苑名家访谈录》。所有这些选题,都是图书馆员思考讨论的所在。相信这些务实的专业论著的出版,对图书馆现时的工作,对图书馆事业未来的发展,一定会有所助益。

在《21 世纪图书馆学丛书》第三辑出版之时,我更加缅怀敬爱的任继愈先生。

丘东江

2009 年 12 月于北京

前 言

数字图书馆的建设在我国已经走过了十多年的历程，它已经从理念、技术的研讨转变为现实的行动。它从一个纯技术领域的问题发展成为了全球性的经济、文化和政治问题，如世界数字图书馆的开通，它的影响和作用已经跨越了资源和技术的范畴。联合国教科文组织总干事松浦晃一郎称：“世界数字图书馆为信息的自由流动、世界团结、文化多样性，以及建立一个兼收并蓄的知识社会提供了宝贵平台。类似数字图书馆这样的项目促使数字技术发挥了应有的文化和社会潜能。”数字图书馆也为现代图书馆的演进注入了新的动力。曾经，我们为数字图书馆概念刚刚出现时而欢呼，数字图书馆作为图书馆发展的方向为众人所向往。而今，它更多的是作为图书馆的一种新的形态为公众所理解和接受。数字图书馆的前景和作用已经越来越得到社会的关注和重视。如何用技术、资源和需求来构建和实现这个目标是图书馆专业人员或相关领域人员需要讨论和研究的问题。

本书避免对该领域概念和理论的探索、总结和归纳，而以建设数字图书馆系统为重点，以可操作为原则，在目前的信息技术条件下，围绕数字图书馆系统的规划、系统构架、资源配置、版权保护、标准规范、系统安全、运营

管理、绩效评估等环节，充分分析数字图书馆建设的具体做法。运用当前数字图书馆的理论，根据数字图书馆使用对象实际功能需求，对系统的建设给出具体的策略。

参与本书编写的都是在图书馆信息技术方面具有丰富实践经验和研究能力的图书馆专业人员。他们具有数字图书馆规划和实践的具体经历，为本书的编写提供了理论和技术保障。

本书由刘晓清制定编写大纲，负责全书的统稿，撰写第1章、第2章，宓永迪撰写第4章、第8章，陈晔撰写第7章、第9章，刘柏嵩撰写第5章、第6章，张杰撰写第3章。宓永迪对全书的体例进行了统一。

由于数字图书馆理论和技术的快速发展，加上本书的撰写时间较紧，同时由于作者的水平所限，不能完全体现编写本书的初衷，在内容和体例上存在许多不足，敬请同行和专家批评指正。

刘晓清

2009年6月27日

目 录

第一章 数字图书馆的信息环境	(1)
第一节 信息资源环境.....	(1)
第二节 信息需求环境.....	(3)
第三节 信息技术环境.....	(5)
第四节 信息用户环境.....	(7)
第五节 信息搜索环境.....	(10)
第二章 数字图书馆规划和类型	(14)
第一节 数字图书馆规划分析	(14)
第二节 数字图书馆类型分析	(21)
第三章 数字图书馆数字资源构建	(25)
第一节 数字资源的特点及类型.....	(25)
第二节 数字资源馆藏建设规划.....	(29)
第三节 数字资源的选择.....	(34)
第四节 数字资源的采购.....	(37)
第五节 数字资源建设中应注意的问题.....	(40)
第六节 数字资源配置建议.....	(45)
第四章 数字图书馆服务功能设置	(50)
第一节 数字图书馆服务特点.....	(50)
第二节 数字图书馆资源导航服务.....	(54)
第三节 数字图书馆检索服务.....	(59)
第四节 数字图书馆虚拟参考服务.....	(62)
第五节 数字图书馆主动推送服务.....	(65)
第六节 数字图书馆个性化信息服务.....	(69)

第七节	图书馆 2.0 下的新型服务	(72)
第五章	数字图书馆运维与评估	(77)
第一节	数字图书馆的系统构成	(77)
第二节	数字图书馆的维护	(86)
第三节	数字图书馆的评估	(90)
第六章	数字图书馆的标准和规范	(103)
第一节	数字资源唯一标识符的标准和规范	(103)
第二节	基本元数据标准	(109)
第三节	专门数字对象描述元数据	(113)
第四节	管理元数据规范	(136)
第五节	对象数据建设的标准与规范	(143)
第六节	元数据互操作、传输协议	(148)
第七节	数字资源长期保存的标准规范	(157)
第八节	信息资源组织和描述、复合数字对象描述,资源 集合、知识组织体系描述标准与规范	(161)
第九节	数字资源系统服务的标准规范	(171)
第十节	计算机中文信息处理规范	(180)
第七章	数字图书馆的版权保护	(189)
第一节	数字图书馆馆藏资源建设的知识产权	(189)
第二节	数据库建设版权问题	(192)
第三节	数字图书馆用户服务的知识产权	(196)
第四节	文献传递中的版权问题	(199)
第五节	数字图书馆版权保护措施	(200)
第六节	数字图书馆建设中知识产权案例	(208)
第八章	数字图书馆建设方案	(215)
第一节	IBM 内容管理平台	(215)
第二节	采用开源软件建设数字图书馆	(224)
第三节	基于商品化数据库软件产品建设数字图书馆	(234)
第四节	基于多种开发平台和工具建设数字图书馆	(241)

第一章 数字图书馆的信息环境

分析和描绘数字图书馆的信息环境，可以让我们明确数字图书馆建设的背景信息和数字图书馆的走向。数字图书馆是整个社会信息的一个子集。就像一项系统设计，对需求环境的了解是进行系统设计前的一项重要工作。调查信息环境是为了回答数字图书馆系统设计中要“干什么”、“为谁干”和“怎么干”而提供依据和资料。本章试图从数字图书馆的外围出发，探讨数字图书馆的环境条件和相关关系，试图描绘出数字图书馆的生存状态。

第一节 信息资源环境

数字图书馆的信息资源包括元数据资源、文献信息资源和网络信息资源。数字图书馆所有的活动的关键是要整合这些资源，使集中有序的客观知识利用发挥最大效益。现在我们所处的信息社会对信息的获取比之任何一个社会时期都更丰富、更方便。现在不是我们去寻找信息，而是被太多的信息所包围。除了传统的传播模式，更多的信息传播模式不断被发明出来。

据因特网数据中心（IDC）研究公司研究表明，在未来三年内数字信息的数量将迅猛增长。这项名为《数字宇宙膨胀：到 2010 年前全球信息增长预测》的研究预计，在未来的三年里，富媒体应用的增长将使得全球数据总量增加到 9 880 亿 GB 以上，这个数量相当于目前所有书籍所含信息量的三倍。IDC 公司的研究还指出，2010 年之前创造出来的数字信息中有 70% 将是被个人创造的。例

如, IDC 公司预计全球数码相机用户在 2010 年之前将拍摄出 5 000 亿张数码照片。如今, 多媒体技术的普及导致非结构化数据如音乐和视频等的数量出现爆炸式增长。虽然 30 多年前的一个普通企业用户文件也许表现为数据库中一排数字, 但是如今的类似普通文件可能包含许多数字化图片和文件的影像或者数字化录音内容。现在, 95% 以上的数字信息都是非结构化数据。在各组织和企业中, 非结构化数据占到了所有信息数据总量的 80% 以上。

在 2003 年的时候, 有专家指出, 如果把信息服务业和国家信息情报部门、科研部门建设的数字化资源都考虑在内, 我国互联网上真正属于科技、经济、文化、教育等方面的中文全文文献, 包括期刊、博士和硕士论文、会议论文、专利、报纸、年鉴等在内, 大约不超过 1 500 万篇, 约占现有印刷文献的 $\frac{2}{5}$; 解决了版权问题上网的图书不超过 15 万种, 约占国内图书的 $\frac{1}{20}$ 。数字化知识资源仅占资源总量的约 $\frac{1}{4}$ 。经过 5 年的发展, 这一现状有了巨大的变化。

信息资源环境应该包含两方面的含义。

一是为了构建数字图书馆信息资源而形成的信息资源组织、加工、整合、传送、检索、安全等各个方面的一系列相关的政策、规则和技术原则、标准规范等。在这些政策、规则和技术原则、标准规范的环境之下, 使数字图书馆信息资源能充分发挥其社会职能和应有效益。这些环境因素随着数字图书馆技术和应用的发展逐渐成熟、完善, 有利于信息资源的构建。

二是数字图书馆信息资源与整个社会信息资源的相关关系。社会信息资源种类繁多, 狭义的信息资源等同于知识、资料和信息, 是信息本身或信息的集合。广义的信息资源认为信息资源是一个贯穿于人类社会信息活动中的, 从事信息生产、分配、交换、流通、消费的全过程的多要素集合。从信息资源角度来看, 图书、报纸、期刊是传统信息资源, 而广播、电视、互联网是现代信息资源。因此, 数字图书馆作为一个系统必须对所处理的信息资源作一个相对

的限定,并协调好与社会信息资源的关系。一般来讲,传统图书馆的资源当然可以转化为数字图书馆的信息资源,在此基础上扩展到现代信息资源的部分。

第二节 信息需求环境

信息需求是信息社会的驱动力,对于信息社会的进步具有重要的意义。宏观的信息需求环境描述可以从年度的中国互联网发展状况统计报告中得到有益的启示。我们从“第23次中国互联网发展状况统计报告”^[3]提取几个关键词可以略见一斑。与信息需求有关联的关键词有:网民规模;互联网基础资源;网民生活形态;网络应用等。

1. 网民规模

截至2008年12月31日,中国网民规模达到2.98亿人,普及率达到22.6%,超过全球平均水平;网民规模较2007年增长8800万人,年增长率为41.9%。中国网民规模依然保持快速增长之势。宽带网民规模达到2.7亿人,占网民总体的90.6%。手机上网网民规模达到11760万人,较2007年增长了133%。农村网民规模增长迅速,网民规模达到8460万人,较2007年增长60.8%,增速远远超过城镇(35.6%)。网民规模的快速增长导致了信息需求的增长。农村网民规模的增长也导致了农村信息需求的增长。

2. 互联网基础资源

2008年中国的互联网基础资源继续保持快速增长,除了IPv4互通技术之外,其他资源的增速基本与网民增长等速或超过网民的增速。IPv4地址的增速已经连续两年落后于网民的增速,IPv4地址的增速如果持续落后于中国网民的增速,未来将成为制约中国互联网发展的瓶颈因素。CN域名和CN域名下网站增速均超过域名总

体和网站总体的增长速度，成为拉动中国互联网资源增长的重要力量。网络基础资源的稳步和快速增长，为信息需求提供了良好的环境。

3. 网民生活形态

网民中的重度用户对互联网作为生活助手的价值认可度高，同时他们对互联网的信任与安全性认可度也较高；另外不可忽视的一点是：越是重度用户，对互联网可能产生的社会隔离认同度越高。

4. 网络应用

互联网的各种应用大致分为如下几类：网络媒体、互联网信息检索、网络通讯、网络社区、网络娱乐、电子商务、网络金融等应用。作为用户自创内容的重要应用，博客自诞生以来，一直保持快速增长的势头，截至 2008 年底，中国博客作者已经达到 16 200 万人。网络游戏在各个应用中排在第六位。在中小学生的应用排序中是第三的位置，网络游戏成为中小学生上网的一个重要应用。大学生使用的前三种网络应用是：网络音乐、即时通信、网络新闻。和总体相比，网络新闻在大学生的使用序列中下降了一个位次，而网络视频则较总体提升了一个位次，排在各个应用的第四位。

信息化的推动力进一步提升了信息需求，第 23 次中国互联网发展状况统计报告认为网民的快速增长有 5 大因素：我国经济的快速发展是互联网用户规模快速增长的基础；国家制订并发布了《2006—2020 年国家信息化发展战略》，《国民经济和社会发展信息化“十一五”规划》等一系列政策，信息化正在成为促进科学发展的重要手段；通信和网络技术向宽带、移动、融合方向发展，数据通信正在逐步取代语音通信成为通信领域的主流；互联网具有高瞻性和高传播性；网民规模的扩张推动网络价值的提升，而网络价值的提升又进一步增强其扩张力。

第三节 信息技术环境

数字图书馆是在计算机技术快速发展的基础上的一种应用和创新。普遍认为建设数字图书馆需要三个方面的要素：资源、技术和服务，这三个要素都有赖于信息技术的进步。信息技术相对数字图书馆来说是一个大背景，我们需要了解这三个要素建立在什么样的信息技术之上。就拿服务要素来说，其工具已从原来的台式计算机扩展到笔记本电脑、上网本（是比全尺寸笔记本电脑更廉价、更轻便的小型电脑）、手机等。以下对数字图书馆的普遍描述可以让我们理出一个数字图书馆与信息技术相关的不可分离的紧密联系。“数字图书馆是一种基于网络环境下的数字化文献信息共建共享的社会信息系统，它集成了计算机、网络、通信、信息处理等多种新技术与新手段，将分散于各种载体、不同地理位置的文献和信息资源进行收集、组织、管理，实现一体化、数字化方式存取，以网络化的方式互相连接，为信息用户提供不受时空和地域限制的网络信息服务。它是一个面向对象的分布式的网状结构模式，通过分布式的信息管理实现全球范围的信息资源共享。”

我们回顾 20 世纪信息技术发展的历史，第三次信息技术浪潮的波涛还在不断地向我们涌来。这波浪潮中我们可以看到自动化浪潮，互联网浪潮到现在的 Web2.0 浪潮，每一波浪潮都是一次信息技术的提升，同时对图书馆的信息处理手段都带来质的飞跃。

自动化浪潮将图书馆传统的信息处理手段提升到了一个计算机处理的时代，在信息技术的浪潮尚处于萌动时期，图书馆没有故步自封，而以极大的热情拥抱了新兴的信息技术。1962 年，美国加利福尼亚大学和南伊利诺伊大学图书馆，分别利用计算机进行期刊和流通管理，开启了图书馆领域计算机应用的先河。1964 年，美国国会图书馆开始 MARCII 试验计划，1967 年发布 MARCII，迈开了图

书馆信息管理标准自动化的步伐。随着计算机联机技术的发展,美国俄亥俄学院图书馆成立了 OCLC (俄亥俄学院图书馆中心/联机图书馆中心)。到了 20 世纪 70 年代,形成了众多的联机检索商用数据库,如 DIALOG 系统、ORBIT 系统、MEDLARS 系统,形成了高度集中的情报检索系统。这些系统的形成和应用基本与信息技术发展的重点——集中式技术模式相适应。在我国,也出现了全国联合编目中心、中国高等教育文献保障系统 (CALIS) 等。各类型图书馆基本完成了从传统到自动化管理系统的转变。

互联网的浪潮带来了更多信息资源的增长和获取信息的便利,推动着数字图书馆的诞生。互联网的初期,网络的信息并不丰富,被誉为信息高速公路的互联网常有有路无车的感慨。因此,图书馆馆藏数字化就成为当务之急,一为提供不受时空限制的查询、存取,从而提高图书馆服务质量,二为保存文献内容。馆藏数字化建设对数字图书馆的推动作用是非常明显的。威廉·阿姆斯 (William Y. Arms) 认为数字图书馆可非正式地定义为有组织的信息馆藏及其相关服务,信息以数字化形式保存,并通过网络进行访问。开始于 1989 年的美国记忆和美国国家数字图书馆项目就是为了提供数字化的馆藏网络服务和资源的长久保存。从国会图书馆数字化发展战略的制定更进一步提出了互联网时代国会图书馆的数字化发展策略。另外,由于互联网主机和网民的大量增加,带来了信息资源的飞速增长,网络上提供和发布的信息可以海量著称,如何有效管理网络上的信息资源成为互联网时代的重大课题。数字图书馆在互联网的浪潮下应运而生。美国数字图书馆联盟 (Digital Library Federation, DLF) 于 1998 年提出的数字图书馆定义认为:数字图书馆是一个拥有专业人员等相关资源的组织,该组织对数字式资源进行挑选、组织,提供智能化存取、解释、传播、保持其完整性和永存性等工作,从而使得这些数字资源能够快速且经济地被特定的用户或群体所利用。随着信息技术的发展,数字图书馆逐渐地由概念演变为现实的行动,数字图书馆的核心问题已成为解决分布式环境下

对数字对象的组织、访问和服务的问题。从两种不同途径发展起来的数字图书馆最终走向了融合和统一。

Web2.0 的浪潮为个人提供了建立个人信息结构与外部信息的链接工具。进入 21 世纪,以 Web2.0 为核心的 Blog、Wiki、RSS、SNS 等网络信息服务大行其道,逐渐为用户所接受,成为当前网络信息服务的主流,这也为数字图书馆在资源整合的基础上如何提供服务手段提供了新的方向。

以上我们概括地描述了技术环境与图书馆、数字图书馆发展的演变关系,显示了技术在其中的作用,有时候技术的发展甚至是决定性的。上海图书馆吴建中馆长说过:“这些年来,我们是被技术推着走的,现代技术不仅影响图书馆赖以生存的社会环境,而且还强行改变着组成图书馆的各个要素。”

第四节 信息用户环境

众多数字图书馆的建设和研究资料较少涉及数字图书馆的用户。图书馆用户的兴趣和水平千差万别。图书馆是休闲读物的来源,也是地区的文化交流中心和信息资料中心。各类人员都可以从本地图书馆获得参考资料。除了因人而异,同一个人在不同时候也有不同的需求,有相同需求的人也会用不同的方法使用图书馆。因此数字图书馆的设计者一定不能假设存在一种统一、特定的使用模式,也不能根据这种模式来开发系统。因此,我们从以下几个方面来分析信息用户环境相关的因素。

一、我国信息用户的基本数据 (截至 2008 年底)

中国网民规模达到 2.98 亿人,较 2007 年增长 41.9%,互联网普及率达到 22.6%,略高于全球平均水平(21.9%)。继 2008 年 6 月中国网民规模超过美国,成为全球第一之后,中国的互联网普及

再次实现飞跃，赶上并超过了全球平均水平。使用手机上网的网民达到 1.176 亿人，较 2007 年增长 1 倍多。中国农村网民规模达到 8 460 万人，较 2007 年增长 3 190 万，增长率超过 60%。网民中乡村人口所占比重不断提升，互联网正在不断向农村地区渗透。与 2007 年相比，网民中大专及以上学历人口进一步下降，高中、初中学历所占比重继续提升。互联网日益向低学历人口普及。

网民的最大构成群体是学生，学生群体的大量存在，一方面极大地活跃着中国的互联网应用，另一方面也降低了中国互联网的商业价值。除了学生之外，党政机关事业单位工作者、企业公司管理者、职员、专业技术人员等文职人员占有较大比重，而占中国人口最大比重的农民、产业服务业工人在网民中所占比重还比较低；与 2007 年相比，网民中无业人群从 11.9% 下降到 5.5%。

信息用户的数量与网民的数量虽然并非完全一致，但存在着正相关的关系。我们可以理解人们通常需要信息是为了两个原因：一是解决问题，二是让生活更精彩。因此，网民的基数预示着一个潜在的庞大信息用户群体。

二、通过网络访问图书馆的用户

有资料显示，2007 年有超过一半的美国人访问了图书馆的资料，他们中许多人通过计算机进行了在线访问而不是直接阅读资料。据美国互联网调查机构皮尤网络与美国生活项目调查数据显示，53% 的美国成年人表示他们在 2007 年访问了图书馆，其中人数最多的是 18 岁至 30 岁年龄段的年轻成年人。这些钟情技术的群体被称为 Y 世代（Generation Y）。调查显示，与非互联网用户相比，光顾图书馆的互联网用户超过了 2 倍。在所有年龄段的美国人中，超过三分之二的人表示他们利用计算机在线访问图书馆的资料。65% 的人用计算机查找信息，62% 的人使用计算机核查图书馆的资源。调研机构表示，目前公共图书馆提供了虚拟家庭帮助和专门的游戏软件，一些图书管理员甚至在虚拟世界第二生命中创作了