

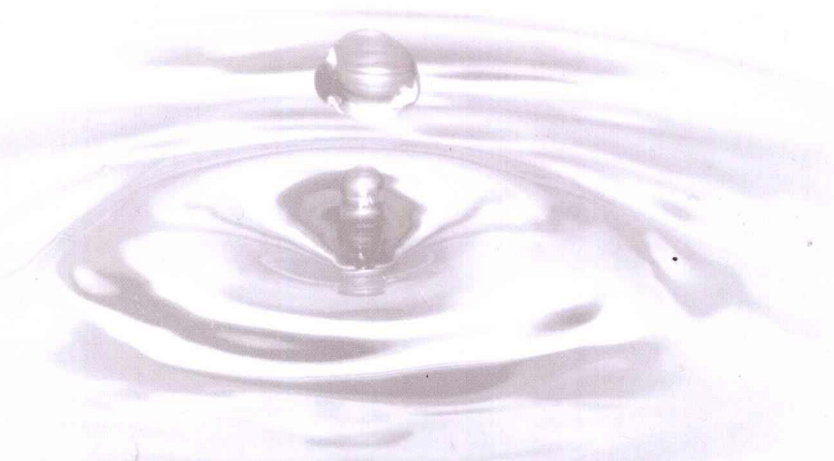
精·品·课·程·立·体·化·教·材·系·列



普通高等教育“十一五”国家级规划教材

数字图书馆导论

夏立新 黄晓斌 金 燕 等 编著



科学出版社
www.sciencep.com

普通高等教育“十一五”国家级规划教材

数字图书馆导论

夏立新 黄晓斌 金 燕 等 编著

科学出版社

北 京

内 容 简 介

本书是教育部确定的普通高等教育“十一五”国家级规划教材。本书系统介绍了数字图书馆的体系结构、设计与开发、“馆藏”发展与维护、信息组织、标准与互操作、用户接口、检索与利用、建设的现实问题。通过本书内容的系统学习,读者能够较系统地了解 and 掌握数字图书馆方面的基本知识、基本理论和基本技能,能够从事数字图书馆的规划、设计、建设和管理的具体实务。

作为一本数字图书馆方面的著作,本书内容全面、翔实,适合作为信息管理与信息系统专业、图书馆学专业的教材,同时也适合作为因特网环境中从事信息资源开发与利用的实际工作者的参考书。

图书在版编目(CIP)数据

数字图书馆导论/夏立新,黄晓斌,金燕等编著. —北京:科学出版社, 2009

(普通高等教育“十一五”国家级规划教材)

ISBN 978-7-03-023536-7

I. 数… II. 夏… III. 数字图书馆-高等学校-教材 IV. G250.76

中国版本图书馆 CIP 数据核字 (2008) 第 209970 号

责任编辑:张 兰 卜 新 / 责任校对:李奕莹

责任印制:张克忠 / 封面设计:耕者设计工作室

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>

西 源 印 刷 厂 印 刷

科学出版社发行 各地新华书店经销

*

2009 年 3 月第 一 版 开本: B5 (720×1000)

2009 年 3 月第一次印刷 印张: 21

印数: 1—3 000 字数: 402 000

定价: 34.00 元

(如有印装质量问题,我社负责调换〈路通〉)

前言

数字图书馆是采用现代信息技术支持的数字信息资源系统，是下一代因特网信息资源的管理模式，正在从根本上改变目前因特网信息分散、不便使用的现状。作为知识经济的重要载体，数字图书馆是国家信息基础设施建设不可缺少的重要内容，也是新形势下国际资源争夺的一个新领域。为此，世界各国高度重视并大力推进数字图书馆建设，一批有影响的实验性项目和工程建设项目纷纷涌现。目前，数字图书馆已成为评价一个国家信息基础水平的重要标志，是 21 世纪各国文化科技竞争的焦点之一。在我国学术界，伴随着数字图书馆概念的研究、国外数字图书馆理论研究和工程建设成果的翻译与评价、国内的技术攻关和工程实施的不断深入，涌现了不少优秀的学术成果。本书是我们多年来开展数字图书馆相关问题研究以及近几年从事数字图书馆课程教学的经验总结。

数字图书馆实践的蓬勃发展急需教育界培养和造就具备数字图书馆建设和管理专业素质的人才。为此，本书的撰写紧密结合了教学改革和课程建设的实际，努力反映数字图书馆在理论研究和实践发展方面的最新成果，体现内容体系的创新。与同类著作相比，本书的特色主要体现在：①从信息管理与信息系统的专业视角来研究和介绍数字图书馆，把数字图书馆看成因特网环境下履行数字化信息资源的采集、加工、组织、存储、检索和利用的信息系统，并以此为主线展开章节的安排和内容的介绍，力求系统介绍数字图书馆涉及的基本知识、基本理论和基本技能，在内容的安排上注重基础性与创新性、系统性与选择性、经典性与前沿性的有机结合。②注重理论和实践的结合。在从理论归纳总结的同时，结合典型的数字图书馆实例，介绍数字图书馆的体系结构、设计与开发工具。通过对本书内容的系统学习，读者既可以系统地掌握数字图书馆方面的理论和专门知识，又可以具备从事数字图书馆规划、设计、建设和管理的动手能力和实践能力。

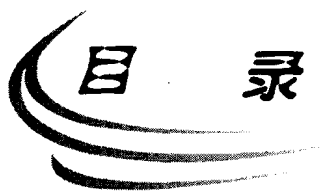
本书的顺利出版,首先要感谢本书的合作者:中山大学资讯管理系黄晓斌教授、郑州大学信息管理系金燕副教授、中国人民大学信息资源管理学院王应解博士、武汉大学图书馆宋登汉博士、陕西理工大学历史文化系方志硕士、武汉生物工程学院信息管理系王平硕士、桂林航天工业高等专科学校图书馆吴玉萍硕士。本书撰写分工如下:夏立新负责本书写作大纲的设计和统稿,并撰写第1、4章,王平、夏立新合作撰写第2章,王应解撰写第3章,宋登汉撰写第5章,黄晓斌撰写第6、8章,方志撰写第7章,金燕、吴玉萍合作撰写第9章。代凤明、陈晨、李小敏、金晶等同学承担了本书部分章节的文字校对工作。没有我们这个团队成员的精诚合作,本书难以顺利完成。本书的出版也得到了教育部新世纪优秀人才支持计划(项目编号:NCET-08-0788)的资助。

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,参考和引用了大量相关文献及网上在线资料。作者谨向这些原文作者以及所有关心和支持本书的撰写与出版的同志表示由衷的感谢!

随着数字图书馆理论研究的不断深入,工程建设项目不断涌现,本书需要补充和完善的内容很多,加之作者水平有限,难免出现疏漏和错误,恳请专家和读者批评指正。我们期待着有机会就数字图书馆理论研究与工程实践中的有关问题与有志于数字图书馆领域研究的同仁进行交流、合作。

编 者

2008年10月



前言

第1章

数字图书馆概述	1
1.1 数字图书馆的产生背景	1
1.2 数字图书馆的相关概念	7
1.3 数字图书馆的特征	16
1.4 数字图书馆的功能	19
1.5 数字图书馆与传统图书馆的业务比较	24

第2章

数字图书馆的体系结构	28
2.1 数字图书馆的基本体系结构	28
2.2 数字图书馆的信息体系结构	30
2.3 数字图书馆的技术体系结构	37
2.4 数字图书馆的系统体系结构	42
2.5 数字图书馆体系结构比较研究	47

第3章

数字图书馆系统的设计与开发	62
3.1 系统开发的基本知识	62

3.2	数字图书馆系统的构成	66
3.3	数字资源整合管理系统	72
3.4	Z39.50 协议在数字图书馆中的技术实现	74
3.5	数字图书馆开发平台	80

第4章

	数字图书馆的“馆藏”发展与维护	88
4.1	演进中的“馆藏”概念	88
4.2	数字图书馆“馆藏”的创建	93
4.3	数字图书馆“馆藏”发展与维护	103
4.4	数字图书馆“馆藏”评价	110

第5章

	数字图书馆的信息组织	117
5.1	概述	117
5.2	数字图书馆信息组织方法和技术	120
5.3	数字图书馆信息组织的具体应用: MARC、元数据、XML/RDF	126

第6章

	数字图书馆标准与互操作	168
6.1	标准概述	168
6.2	数字图书馆标准规范的基本问题	173
6.3	数字图书馆标准规范体系	191
6.4	数字图书馆的互操作	202

第7章

	用户接口	211
7.1	用户接口理论概述	211
7.2	用户接口设计	218
7.3	用户接口评价	226

7.4 用户接口的发展	235
-------------------	-----

第8章

数字图书馆的检索与利用	241
8.1 数字图书馆门户	241
8.2 数字图书馆的跨库检索	248
8.3 跨语言信息检索技术	261
8.4 基于内容的检索技术	271
8.5 数字图书馆的知识检索	275
8.6 信息检索的可视化	287

第9章

数字图书馆建设的现实问题	297
9.1 经费与运作模式问题	297
9.2 知识产权保护问题	303
9.3 标准化问题	311
9.4 技术问题	317
9.5 其他问题	321

主要参考文献	325
--------------	-----



数字图书馆概述

本章介绍数字图书馆的产生背景及其发展历程，从不同角度对数字图书馆的概念加以理解，并将数字图书馆概念与人们常见的虚拟图书馆、复合图书馆、电子图书馆等概念加以比较，进一步归纳出数字图书馆的 5 个特点，介绍了数字图书馆的功能以及数字图书馆在业务功能上与传统图书馆的不同，有利于读者对数字图书馆形成一个大体认识。

1.1 数字图书馆的产生背景

20 世纪末，随着计算机技术、数字存储技术、现代通信技术、网络技术的飞速发展与有机结合，数字化信息广泛存在于人们身边。像其他格式的信息一样，数字信息在整个生命周期内运动，它们被产生、编辑、描述、标引、传播、获得、使用、注释、修订和再产生。而当我们谈论图书馆时，首先想到的是雄伟的建筑、汗牛充栋的书库和宽敞明亮的阅览室。但在数字化信息的冲击下，一种新型图书馆——数字图书馆应运而生，它打破了人们对图书馆传统认识的时空观念，在基于计算机网络特别是因特网的多媒体信息管理系统下，数字图书馆的信息不再存放在一栋物理建筑里，而是分布在世界各地的网络系统中，呈现出一种虚拟状态。

从社会、经济、技术发展大背景的角度来分析数字图书馆产生与发展的影响因素，可以进一步加深对数字图书馆的了解，真正认识其存在与发展的本质。

1.1.1 数字图书馆的发展历程

数字图书馆是现代信息技术发展与人类信息知识需求发展相结合的产物。人

类社会中任何一种事物的产生与发展都与人类生产、生活的需求分不开。

人类文明绝大多数是通过文献记录下来并流传后世的。图书馆作为文献最重要的保存、利用场所之一,通过开发利用文献,传播信息和知识,积累、继承并创造新的知识,成了人类文明发展的重要推动者。随着现代信息技术在文献和图书馆中的应用,这种推动作用将越来越大。

回顾 20 世纪的 100 年,影响人类社会各个方面程度最大、速度最快、震撼最强的,莫过于至今仍在飞速发展的信息技术。其影响范围涉及社会生活的各个方面,图书馆也不例外。从发展历程上看,信息技术的发展水平及其在图书馆领域的应用水平和程度,在很大程度上被视为划分图书馆发展阶段的重要依据。

从总体看,20 世纪 60 年代以前的图书馆是以收藏印刷品为主,通过卡片式或书本式目录反映馆藏,通过检索刊物反映本馆可能没有收藏但已确实存在的文献,图书馆业务工作的开展基本上以手工为主。

20 世纪 60 年代以后,以计算机、通信以及网络为核心的现代信息技术开始全面渗入图书馆领域,将图书馆推向自动化发展阶段。1960 年,美国人尼诺提出的编印目录卡片的机械化装置,被认为是最初的机读目录。1962 年,美国加利福尼亚大学和南伊利诺伊大学分别应用计算机进行期刊管理和流通管理。1964 年,美国国家医学图书馆用计算机编制《医学文摘》,并生产 MEDLARS 数据库磁带。1966 年,美国国会图书馆开始了机读目录计划(Machine Readable Catalogue, MARC),特别是 1969 年开始正式发行 MARC II 机读目录,为实现图书馆大部分业务工作的自动化奠定了基础,为图书馆资源共享创造了条件。

20 世纪 70 年代,计算机与通信技术相结合,促进了图书馆联机系统和网络化的发展,出现了一批联机编目网络和一些商业性联机检索系统。例如,OCLC(Ohio College Library Centre)译为“联机计算机图书馆中心”(Online Computer Library Centre),已发展成为全球最大的书目数据库中心;Dialog、ORBIT、MEDLARS 等联机检索系统先后建成并投入商业运行。

20 世纪 80 年代,图书馆自动化的显著特点是图书馆联机系统进一步发展,与各馆建立的自动化集成系统连成网络。各馆普遍实现了采购、编目、流通、连续出版物管理、检索、馆际互借以及行政管理等方面的自动化,建立了机读目录数据库和二次文献数据库。用户通过联机公共存储目录(OPAC)查询馆藏书目,通过联机检索终端查找二次文献信息。

从图书馆采用现代信息技术的历程看,图书馆一直强调内部业务工作的自动化和检索工具编制与使用的自动化。但由于受到计算机运算速度和存储容量等的限制,图书馆当时难以实现对文献本身存储与检索的计算机化。

到 20 世纪 90 年代,由于高密度存储技术、多媒体技术、通信技术、网络技术现代信息技术的高度发展、不断突破及有机结合,图书馆已可以对文献本身

进行管理。从此,图书馆逐步迈进新的阶段——数字图书馆。

而图书馆所收藏的文献从产生以来,也一直根据社会需求的发展而不断地应用新的信息技术进行保存与管理。从最早的山石、金属器皿、甲骨,到后来的竹简、丝帛,再到后来的纸张,20世纪后有了唱片、录音带、录像带,特别是产生数字技术后有了CD-ROM光盘、DVD光盘等,包括网络与数据库管理以及可以预见到的数字图书馆系统等。可见文献资源的生产与管理总是走在新技术应用的前沿。进入21世纪,数字化的文献资源已成为社会需求与应用的主体。相应地,文献资源数字化生产与管理技术及其解决方案便成为重要的发展战略性课题。

社会公众对于图书馆的要求,主要是从馆藏文献中获取系统的、成型的信息——知识。即便是在欧美经济文化最为发达的国家和地区,公众的需求也主要表现为系统知识的获取。大量实际调查与研究表明,对科研人员、企业决策人员、行政管理人员等来说,处于当代世界的激烈竞争氛围中,第一时间获取最新信息自然是非常重要的。但是对社会的绝大多数成员来说,情况却明显不同。社会环境及其个人生活的种种变化,随时可能要求他们涉足一个新的领域,此时他们特别需要的是这一领域最新的动态和相关知识。因此,图书馆将是他们获取不同领域知识的最佳场所。图书馆已开始成为终身学习环境的营造者。

进入新世纪,置身于人类社会全面信息化、知识化趋势下的文献资源生产、管理与利用,必然以数字化、网络化为主要特征,以方便、高效地满足人们知识获取与学习需要为目标。因此,新技术的发展与人类信息、知识需求发展相结合的结果,必然产生数字图书馆这一新型信息与知识的处理系统。可以说,数字图书馆就是使记录人类文明的载体——文献在最新的相关技术应用于其中后,能够满足人类对知识、信息最新需要。

1.1.2 数字图书馆发展的影响因素

从技术、需求方式、经济等具体角度讲,数字图书馆的产生及其持续发展曾经并仍将受以下几点主要因素的影响。

1. 新战略资源的无限增长与有效利用的矛盾亟须解决

自20世纪60年代以来,人类在生产实践中逐渐认识到,资源不仅有各种物质形态,也有知识、经验、技术等非物质的信息形态。前者包括物质资源和能源资源,后者可称为第三资源即信息资源。信息资源作为一种战略资源,现已成为知识经济时代决定社会发展进步的决定性力量。但在现代经济活动的发展为信息的生成、传递、存储和积累提供了条件的同时,各种形态的信息以突飞猛进的速度增长并迅速积累起来,很快就达到了一个非常庞大的基数。据统计,20世纪40年代以来所产生和积累的信息量超过了此前人类所有的信息量之和。根据

《自然》杂志 1999 年 8 月发表的研究结果, WWW 有近乎 80 亿页可公开存取的信息^①; WEB 的规模仍在呈指数增长。根据另一项估算, 1998~2002 年, WEB 的数量增长了 3 倍^②。由此可见, 网上信息资源的发展势头已远远超过当年引发“情报危机”时文献的增长规模, 正在出现因特网环境下的“信息爆炸”。

如此海量的信息增长与有效的信息管理和获取、利用两者间必然存在着尖锐的矛盾。随着信息量的增加、用户需求水平的提高, 矛盾的尖锐程度几乎无法调和。即使使用最强大的搜索引擎, 当用户提出一个具体问题时, 返回的结果仍是大量且繁杂无序的。这样, 用户只能在众多的网站和网页中换来换去, 浪费了时间不说, 真正准确、需要的结果却没有得到。由此可见, 因特网神奇地通过数以万计的计算机连接, 把整个地球联系在一起, 人们在学习、生活和工作中享受便利的同时, 网上信息的急剧增加与用户对信息的个性化需求之间的矛盾不但没有得到解决, 反倒愈加突出, 人们似乎又陷入“情报爆炸”时代所面临的困境。虽然因特网使我们可以得到的信息比以前任何时候都多, 但并不意味着我们总能得到我们想要的信息。

信息利用者, 特别是知识型信息的利用者, 已成为信息海洋中无助的孤独者, 迫切需要一个真正的知识导航系统。毫无疑问, 这个真正的知识导航系统应该就是数字图书馆。

2. 科学技术的发展为数字图书馆提供了实现手段

数字图书馆是一个宽带多媒体网络和海量信息管理系统, 它所面对的管理对象、存储对象和技术领域远远超出了传统图书馆的范围。存储对象包括数字化的图书、音像、新闻、美术作品、照片、雕塑、电影、软件、电子出版物、因特网内容、卫星数据、气象数据、地理数据、政府文件等各种各样的人文与科学数据, 因而牵涉大量的技术性管理问题。计算机技术、存储技术、网络技术、信息处理技术的不断发展, 使得计算机性能不断提高。根据摩尔定律^③, 计算机计算能力每 18 个月翻一番, 即每 5 年增长 10 倍。但现代信息技术的发展已突破了摩尔定律, 如计算机的三维图形性能、广域网的带宽、光纤传输容量、存储器的容量、信息处理技术、智能处理技术等都在以超常的速度发展, 特别是这些技术在应用领域的广泛普及, 为数字图书馆的建设与发展提供了现实技术条件和应用条件。

3. 人们需要一个真正的文献信息资源集成共享系统

对通过因特网来查找图书馆藏书和专题数据库资料的读者来说, 他们目前还

① Lawrence S, Giles C L. Accessibility of information on the web. Nature, 1999, 400: 107

② OCLC Office of Research. Size and growth statistics. Web Characterization Project. <http://www.oclc.org/research/projects/archive/wcp/stats/size.htm>. 2008-7-9

③ <http://baike.baidu.com/view/17904.htm>

需要关心从哪里得到所需的内容,需要到不同的图书馆数据库中查找藏书信息,并不断地做出判断和进行烦琐的操作。或者用较多的时间等待自己成为新一家图书馆数据库的登记用户,以便使用各馆并不免费提供的内容。目前,图书馆系统内普遍存在着馆藏资源不能共享、文献或数据库严重重复或缺藏的问题。

事实上,规模再大的图书馆也无法将世界上所有的出版物收集齐全,而且也无法包揽所有文献信息的处理工作,更无法满足众多个性化的服务需求。在信息资源无限扩大的今天,如果一个图书馆不借用外力、不共享已有社会劳动成果,而只是一味地闷头工作的话,是难以完成自己使命的。因此无论是从读者需求的角度,还是从图书馆业务管理工作的角度,通过数字图书馆系统来实现各馆传统文献和数字化文献资源的共享都是一个必然趋势。而且,这种共享系统将逐步扩展到图书馆领域之外,由图书馆与包括档案馆、出版社、新闻媒体、博物馆在内的资源拥有者共同组成。因此,数字图书馆系统必然是一个资源集成管理系统。这个系统既要管理传统文献,又要管理与传统文献在内容上紧密联系的数字化文献,而且不分类型,不分内容。只有这样,才能实现真正意义上的资源共享。

4. 人们需要一个全新的知识型信息的管理与获取平台

信息利用是极易迷失方向的。很多不全面甚至是错误信息被利用的例子屡见不鲜。而现有的因特网系统并未能从根本上解决这一问题。那么如何避免这样的问题呢?建立一个信息过滤系统,甚至是一个知识的生产系统是必然的选择。从前面的分析看,数字图书馆就应该是这样一个信息过滤系统。因此,建立一个可供全人类进行知识的管理、生产与获取利用的平台,应该是数字图书馆建设的目标。当前各地建设数字图书馆的热情很高,有的将一个馆的部分资料数字化,就称自己是数字图书馆;有的公司将一些图书数字化以提供网络阅读,也称建成了数字图书馆;有的甚至建立了一些可进行检索的网页,也称建成了数字图书馆系统。这些都是真的数字图书馆吗?

很多人没有意识到信息处理工作也是容易迷失方向的。数字图书馆作为一种高级的信息处理工作,如何才能不迷失方向呢?明确建设一个知识型信息的管理与获取平台是非常必要的。诺贝尔奖获得者 H. 西蒙在他的名著《决策管理新科学》中认真地分析了 20 世纪 60 年代美国企业一批计算机信息系统失败的现象。其表现是,对这些原打算为企业高层领导服务的系统,领导实际上很少问津。H. 西蒙分析的原因是这些系统把方向搞错了。它们都把目标放在为领导提供更多的详细资料上,而在信息爆炸的社会,领导并不需要更多的资料,他们的信息负荷已经过载了。高层领导真正需要的是筛选信息的系统,而不是增加资料的系统。让我们就这个例子冷静地思考一下,在今天的数字图书馆研究和建设中,领导者和技术人员关注最多的是计算机技术、存储技术、网络技术、海量资源建设、技术标准等,用户的需求研究和资源内容研究明显被忽视。在这两项还没有

清楚时,大量的资源已经被数字化并开始提供服务了。这些系统是一个筛选信息的系统吗?我想今天的用户,就像20世纪60年代美国的企业领导者们一样,最终会做出理智的判断。同样,清醒的数字图书馆设计与建设者,最终也会做出明智的选择。

5. 数字图书馆作为一个新的经济体有其不断发展的内在动力

现在很多人自觉或不自觉地将数字图书馆视为公益型的基础信息设施,这在其建设的初始阶段无疑是正确的。然而,数字图书馆最终必将成为一个不断发展的经济体,这是由其与经济的密切联系所决定的。

首先,数字图书馆的建设需要信息技术厂商广泛参与。无论是计算机设备、存储设备、网络设备、数字化设备、通用软件的应用,还是应用软件开发与资源有偿使用等,都使得数字图书馆的建设引来无限商机。其次,从数字图书馆的自身运行来说,它已经不可能是传统图书馆时代以公益性服务为主的非营利性机构。数字图书馆管理的知识性资源是知识经济时代获取利润的源泉,谁掌握了这些知识性信息,谁就有可能创造出更多的精神和物质财富。因此,数字图书馆的一个主要功能,就是不断地发掘知识信息资源及其相关要素的经济功能,以获取合法的经济收益。再次,数字图书馆可以实现有关文献资源机构的资源共享。这样可以有效解决文献资源(含数字化资源)的重复购置。而新的数字化资源进入系统后,其他机构就可以充分利用这部分资源,从而大量节省人力、物力,减少资金投入,提高工作效率,这也是最大的效益。最后,在资源生产者的知识创造与所有权利真正得到法律的保护,并通过数字图书馆系统得到技术保护和经济收益时,一个知识增值的链条就建立起来了。

知识的创造者会由于数字图书馆系统给其带来经济收益,而愿意提供自己拥有知识产权的作品,从而利于数字图书馆的资源不断丰富发展。而数字图书馆的用户,则因此得到更多真正需要的知识产品。这不正是一个以知识为产品的有效运转的经济体吗?知识创造者从数字图书馆系统中能够获得源源不断的经济收益,这种收益并将随着数字图书馆系统的完善而逐渐增大。

这些经济收益和效益,必然促使数字图书馆作为一个知识经济时代的新型经济体,在市场中不断发展完善,满足社会与经济发展的需要。这也就是世界上数字图书馆建设浪潮为什么会一浪高过一浪、技术创新与建设热度不减,而且数字图书馆管理方绝大部分具有公司背景,有的甚至是由公司来建设的原因了。目前,发展较好的数字图书馆有中国数字图书馆有限责任公司、书生之家数字图书馆、超星数字图书馆等。而且很多成功的信息资源服务公司首先源于其企业化管理背景,如慧聪国际资讯有限公司、万方数据集团公司、清华同方光盘股份有限公司等,而并不在于其原来是否是文献和相关资源的所有者。这说明面向市场经济的体制、运行机制是核心。如果数字图书馆的建设与发展不能深刻地认识到并

解决好这一问题,而是仍沿用传统图书馆的管理体制和运行机制,是不可能得到良性发展的。

1.2 数字图书馆的相关概念

1.2.1 数字图书馆概念的理解与分析

我们首先从数字图书馆基本概念入手,从不同角度看待数字图书馆,以便对其内涵、特点有更深刻的理解。

1. 数字图书馆概念的多种理解

从目前数字图书馆在国内外的发展和研究情况看,数字图书馆的概念涉及众多领域,其中有很大一部分是由图书馆界以外的人士进行界定的,如计算机科学家、软件工程专家、专家系统研究人员、图像处理和语言处理专家、网络专业人员、教育界人士、知识产权专家等。从中可以看出,数字图书馆这一课题实际上是一个与信息技术密切相关的跨学科、多层面的新兴领域。每个领域的探索都从不同的角度丰富了数字图书馆研究的内涵,而且对数字图书馆问题,甚至字面本身都有不同的理解。

对于一般用户而言,数字图书馆只不过意味着传统图书馆的计算机化;对于当代 WWW 的用户而言,效率、组织、功能和可用性日益增强的 3W 就是数字图书馆,从这个意义上讲,3W 是全球最大的数字图书馆;对于计算机专业人员而言,数字图书馆只不过是一个分布式的基于文本和图像的信息系统、分布式信息服务的一个集合,是一个网络化的多媒体信息系统或者一种分布式的互联信息空间;教育界人士则把数字图书馆视为融合正式学习、非正式学习、职业性学习的学习支持手段和工具,甚至认为数字图书馆就是数字化学校,它将专门技巧、专业知识以及综合浏览等进行组装以用于创新性发展、自我指导和非正式学习;而对于图书情报学研究者来说,数字图书馆将以一种新的方式执行图书馆的功能,包含新型的信息资源、新的采访和馆藏发展方式(更多的资源共享和电子订购服务)、新的存储和保存方法、新的分类和索引方式、与用户新的交互模式、对电子系统和网络更多的依赖性以及图书馆在人员智力构成、组织、经济等方面的显著变化;对于信息管理与信息系统专业^①的研究人员来说,数字图书馆只是信息管理系统中的一种,正如工资人事管理系统、银行的电子支付系统、会计中的电算化系统、商务活动中的电子商务系统等分别是信息管理系统中的一种一

^① 信息管理与信息系统专业是1998年教育部在修订高等教育本科专业目录的过程中,在原科技情报专业、信息学、经济信息、管理信息系统等专业的基础上,调整合并设立的新专业。

样,它用于图书馆数字化信息的采集、加工、存储、提供服务等业务工作。

现将有影响的关于数字图书馆概念的描述和提法归纳如下,以期通过共同的探索,得出一个绝大多数人能够接受、对实际工作有指导意义的数字图书馆概念。

1990年,美国密歇根大学的研究人员首次提出了“数字图书馆”的概念,并将其定义为:数字图书馆是若干联合机构(federated structure)的总称,它使人们能够智能地(intellectually)和实实在在地(physically)存取全球网络上以多媒体数字化格式存在的、为数巨大的且仍不断增多的信息。

1995年,一个题为“交互作用,定标及数字图书馆研究议事日程”的数字图书馆专题研讨会给数字图书馆做了如下定义:“一种多媒体数据与把数据表现为信息与知识信息管理方法的两者的有序集合。”一年后,此定义又被修改为:“数字图书馆是一种多媒体数据与信息方法的有序结合,这种管理方法把数据表现为一种对不同社会环境的人们有用的信息知识。”两个定义都指出了数字图书馆能增进馆藏的扩充、组织、存取、注释及保存,且都论及了数字形式的信息以及存储在物理媒体上信息的数字化管理,但学者的定义又加进了一个重点,即需要将用户和用途作为数字图书馆研究的一个重要部分。1997年3月,由美国国家科学基金会(National Science Foundation, NSF)主办的桑塔菲分布知识工作环境计划研讨会又将这个定义做了进一步的扩充,数字图书馆的定义不仅仅是数字化的收益与信息管理工作两者的对应词,它更应说是一种环境,这个环境将收藏、服务方式及人结合起来,以支持数据、信息和知识的搜集、传播、利用及存储。2000年,《D-Lib杂志》的创办者 William Y. Arms 把数字图书馆非正式地定义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。定义的核心在于说明信息是有组织的。从卫星发往地球的数据流不能直接作为图书馆的馆藏,同样的数据一旦经过系统化的组织,便成为数字图书馆的馆藏。

研究者与实践者对数字图书馆提供了越来越多更富创造性的定义,他们不断扩展思路,充分考虑数字图书馆的含义可能延伸的程度,重要的是不能让类似数字图书馆这样一个迅速出现的研究领域过早地“定型”,也就是说要让这个定义长期地尽可能地开放。

俄罗斯多数学者认为,数字图书馆虽与传统图书馆的自动化有许多共同点,然而二者并不等同。索科洛娃和利雅别夫将数字图书馆定义为:数字图书馆是一个分布式信息系统,它能够保存、有效地利用各种各样的电子文献,对于最终用户这些电子文献可以通过全球网络传输而便利地获取。俄罗斯科学院社会科学研究所副所长格鲁霍夫认为,正在日益得到普及的电子图书馆,是汇集广袤信息资源并为用户提供有效利用手段的大规模的综合体。他针对将“电子图书馆”、“虚

拟图书馆”和“数字图书馆”等同的观点发表了不同看法,认为三个要领有所不同:虚拟指网络信息资源;电子则包括现实馆藏;数字化比电子化更具时代意义,数字图书馆是一种具有发展潜力的概念,因而可以代替前二者。

日本图书馆情报大学的田佃孝一认为数字图书馆是指不局限于文本,包含画像、影像、音响等在内的多媒体图书的收集、存储、流通以数字信号的形态统合进行处理的图书馆。但同校的杉本重雄却认为数字图书馆未必含有“馆”的含义,从其构成看,数字图书馆是指将数据资料以数字化方式收集存储,通过网络来利用的机构。日本中央大学的齐藤孝则认为数字图书馆是在像因特网这样的赛柏空间(cyberspace)展开的超文本的集合体。

我国学者对数字图书馆含义的解析中,具有代表性的有以下几种:

汪冰曾经把国外图书情报界对数字图书馆的定义归纳为:计算机网络化系统,新信息技术在图书馆及类似机构中的应用,只存储电子资料的新型图书馆,强调集成信息服务和广泛存取信息的一种概念和思想。他本人也将数字图书馆广义理解为“计算机可处理信息的集合或此类信息的一个存储处”。

天津理工学院图书馆的赵伟 1999 年在《情报科学》上发表题为《数字图书馆研究的历史和现状》一文,提出:数字图书馆是以数字形式存储和处理信息的图书馆,是将计算机技术、通信技术、微电子技术等合而为一的信息服务系统,主要由三大部分构成:数据库管理服务系统、图书馆网络通信系统、数字化的信息资源系统。

吴志荣在《数字图书馆——从理念走向现实》中把数字图书馆定义为:数字图书馆是运用当代信息技术,对数字信息资源进行采集、整理和存储,并提供给所有链接网络的用户,是为一定的社会、政治、经济服务的文化教育机构以及这种机构的组合。

陈敏在《情报学报》1999 年第六期上发表题为《Internet 时代的数字图书馆的建设》的文章,提出:数字图书馆是传统图书馆功能的扩展,它对信息进行搜集、转换、描述,并以计算机可处理的数字化形式存储馆藏数字化信息和网络数字化信息,以智能化的信息检索和统一友好的检索界面,利用先进的信息处理技术和互联的计算机网络,提供多种语言兼容的多媒体远程数字信息的信息服务机构。

徐文伯(中国数字图书馆工程领导小组组长)对于这一概念,有两种提法。其一认为:所谓数字图书馆就是对有高度价值的图像、文本、语音、音响、影像、影视、软件和科学数据等多媒体信息进行收集,进行规范性加工,进行高质量保存和管理,实施知识增殖,并提供在广域网上高速横向跨库链接的电子存取服务。同时还包括知识产权、存取权限、数字安全管理等。它的特点是收藏数字化、操作电脑化、传递网络化、信息存储自由化、资源共享化和结构连接化。其