



数字图书馆 概论

Shu Zi Tu Shu Guan

Gai Lun

编著 / 过仕明
杨晓秋

黑龙江科学技术出版社

数字图书馆概论

过仕明 杨晓秋 编著

黑龙江科学技术出版社

中国·哈尔滨

图书在版编目(CIP)数据

数字图书馆概论/过仕明,杨晓秋编著. -哈尔滨:

黑龙江科学技术出版社,2006.8

ISBN 7-5388-5183-6

I. 数... II. ①过...②杨... III. 数字图书馆概论

IV. G252.7

中国版本图书馆CIP数据核字(2006)第092071号

责任编辑 王嘉英

封面设计 王嘉英

数字图书馆概论

SHU ZI TU SHU GUAN GAI LUN

过仕明 杨晓秋 编著

出版 黑龙江科学技术出版社

(150001 哈尔滨市南岗区建设街41号)

电话(0451)53642106 电传 53642143(发行部)

印刷 哈尔滨市印苑广电印务有限公司

发行 黑龙江科学技术出版社

开本 787×1092 1/32

印张 11.125

字数 244000

版次 2006年8月第一版.2006年8月第一次印刷

印数 1-1000

书号 ISBN 7-5388-5183-6/G·559

定价 22.00元



作者简介：过仕明 男 1959年3月出生 安徽省和县人 汉族 哈尔滨师范大学图书馆馆长 副研究馆员 吉林大学管理学院情报学专业 博士生 黑龙江省图书馆学会常务理事 黑龙江省高校图书情报工作指导委员会副主任 主编著作4部，主持、参加各级各类科研项目11项，在《情报学报》《中国图书馆学报》《大学图书馆学报》《情报资料工作》《情报科学》《图书馆建设》等学术刊物上发表论文20余篇。



作者简介：杨晓秋 女 1973.8.2生 毕业于黑龙江大学信息管理系 学士学位 哈师大图书馆信息咨询部主任 参编、主编信息检索课教材3部；发表学术论文1篇；参与校级课题1项。

内 容 提 要

数字图书馆已成为全球范围内信息基础设施建设的热点领域,是 21 世纪全球文化与科技竞争的焦点之一。本书主要以国内外有关数字图书馆的最新研究成果为基础,概述数字图书馆的相关概念、体系结构、关键技术、标准规范、存储与检索、信息服务和知识产权问题等内容,并预测其未来发展走向。该书特别适合大中型图书馆人员以及与信息检索有关的研究生、本科生阅读,同时也为科研人员、情报人员和数字图书馆爱好者了解数字图书馆相关知识提供了帮助,是学习数字图书馆基本理论和基本知识的一本值得推荐的书。

目 录

第一章 数字图书馆及其相关概念	1
1.1 数字图书馆的产生背景	1
1.2 数字图书馆的定义及特征	4
1.2.1 数字图书馆的定义	4
1.2.2 数字图书馆的特征	7
1.3 数字图书馆与传统图书馆的关系	13
1.4 数字图书馆、电子图书馆和虚拟图书馆的比较分析	17
1.4.1 数字图书馆、电子图书馆和虚拟图书馆名称的 使用情况	17
1.4.2 数字图书馆、电子图书馆和虚拟图书馆之间的 区别	19
第二章 数字图书馆的体系结构	21
2.1 数字图书馆的信息体系结构	22
2.2 数字图书馆的技术体系结构	26
2.3 数字图书馆系统的结构模型	35
2.4 数字图书馆体系结构比较研究	41
2.4.1 基于三层客户机/服务器模式的数字图书馆结构	42
2.4.2 分布式数字图书馆结构	45
2.4.3 基于代理协作的数字图书馆体系结构	51
2.4.4 Interspace 和 CSTR 的体系结构	53
2.4.5 数字图书馆体系结构的评价标准	55
第三章 数字图书馆的关键技术	57

3.1 Web 技术及应用	57
3.1.1 Web 概述	57
3.1.2 Web 体系结构	59
3.1.3 Web 开发技术	66
3.2 中间件技术	75
3.2.1 中间件的基本概念	75
3.2.2 中间件技术简介	76
3.3 数据仓库和数据挖掘技术	88
3.3.1 数据仓库和数据挖掘技术概述	88
3.3.2 数据仓库、数据挖掘技术与数字图书馆	94
3.4 数字图书馆的集成技术	97
3.4.1 系统集成的概念	98
3.4.2 数字图书馆系统集成内容	98
3.4.3 数字图书馆应用集成的方式	100
3.5 数字图书馆系统的安全技术	101
3.5.1 概述	102
3.5.2 影响网络信息安全的主要因素	107
3.5.3 网络安全技术	110
3.5.4 操作系统安全技术预防机制	113
3.5.5 应用服务安全技术预防机制	115
第四章 数字图书馆标准规范的制定和应用	117
4.1 数字图书馆建设的标准规范体系	117
4.2 数字图书馆建设的相关标准规范	127
4.2.1 资源编码标准(Resource Encoding Standards)	128
4.2.2 资源标识标准(Resource Identification Standards)	136
4.2.3 资源描述标准(Resource Description Standards)	140

4.2.4 数据存档标准(Data Archiving Standards)	145
4.2.5 互操作标准/检索服务标准(Interoperability Standards)	146
4.3 国外数字图书馆标准规范的制定和应用	148
4.3.1 美国数字图书馆标准规范的制定和实践	148
4.3.2 英国数字图书馆标准规范的制定和实践	165
4.3.3 加拿大数字图书馆标准规范的制定与实践	170
4.4 国内数字图书馆标准规范的制定和应用	171
第五章 数字图书馆信息存储与检索技术	188
5.1 数据的存储与备份	188
5.2 数据压缩技术	201
5.3 基于内容的信息检索技术	213
5.4 跨语言信息检索	225
第六章 数字图书馆的信息服务	240
6.1 数字图书馆服务的特点	240
6.2 数字参考服务	245
6.3 主动推送服务	255
6.4 定题服务	261
6.5 个性化信息服务	265
6.6 用户培训服务	273
第七章 数字图书馆知识产权保护问题	277
7.1 数字图书馆信息资源建设的知识产权问题	277
7.2 数字图书馆用户服务中的知识产权问题	293
7.3 数字图书馆版权许可使用与版权集体管理制度	308
第八章 数字图书馆的发展与未来	322
8.1 数字图书馆的研究历程	322

8.2 数字图书馆的发展概况	325
8.2.1 国外数字图书馆的发展概况	325
8.2.2 国内数字图书馆的发展概况	329
8.3 数字图书馆的未来	333
8.3.1 国外数字图书馆的发展趋势	333
8.3.2 国内数字图书馆的发展趋势	337
8.3.3 数字图书馆发展存在的问题	339
8.3.4 数字图书馆的未来走向	341
参考文献	345

第一章 数字图书馆及其相关概念

图书馆是文献收集、加工、保管、传播的信息场所。随着计算机技术、尤其是信息技术的飞速发展,使传统图书馆文献存储形式和传播方式发生了重大的变革,其显著特征是文献存储形式的数字化和传播方式的网络化。

20世纪80年代末,“电子图书馆”和“虚拟图书馆”的概念就在发达国家引起重视,并进行了相关的研究和实践,取得了很大的成功。进入到20世纪90年代以后,美国的信息高速公路将图书馆、学校、机关、商业机构、医院、家庭乃至个人连接起来,并对所存储的信息资源提供网络检索,实现了真正意义上的资源共享,由此产生了数字图书馆的雏形。目前,我国数字图书馆的建设已取得显著的成效,读者正在利用网络,在数字图书馆中汲取知识。

1.1 数字图书馆的产生背景

数字图书馆的发源地是美国,并在美国逐渐发展和创建起来,从而带动了世界范围内的推广和发展。到目前为止,美国在数字图书馆建设方面一直处于国际领先地位。

1.数字图书馆构想的提出

1945年,美国著名科学技术管理学家布什(V.Bush)先生在《大西洋月刊》上发表了《诚如我们想象的那样》(As We May Think)一文。文中,他首次提出将传统的图书馆馆藏文献的储存,

查找机制与计算机结合起来,构思并描述了他所设想的一种 Memex 装备机械化的个人文档与图书馆,即台式个人文献工作系统,能存贮书、记录和通讯的装置。1948年,美国数学家维纳(N. Wiener)成为第一个指出电子计算机将能够在图书馆运作中大显身手的学者。1978年,美国著名图书馆学家兰卡斯特(F.W. Lancaster)发表了《走向无纸的时代》和《电子时代的图书馆员》两部论著,第一次向世人详细描述了电子图书馆的前景。1979年,英国不列颠图书馆的哈利(A.J. Harley)提出了“虚拟图书馆”的概念。1988年,美国国家科学基金会(NSF)的伍尔夫(W. Wulf)撰写国际合作白皮书,正式提出了数字图书馆的概念。

2. 数字图书馆的启动与实施

早在1991年,美国俄亥俄州政府做出了启动俄亥俄网的决定,计划投资2500万美元建立州内图书馆网络中心,该网络定名 Ohio LINK。1992年,美国政府在制订“高性能计算与通信”(HPCC)国家攻关项目中,第一次将发展数字图书馆列为“国家级挑战”(National Challenge)项目之一。同年,美国前副总统阿尔·戈尔提出美国信息高速公路法案。1993年9月,美国政府宣布实施一项新的高科技计划——“国家信息基础设施”(National Information Infrastructure,简称NII),旨在借助于因特网(Internet),兴建信息时代的高速公路——“信息高速公路”,使所有的美国人方便地共享海量的信息资源。

紧随美国的信息高速公路计划,世界各国纷纷效仿,相继提出各自的信息高速公路计划,并投入巨资实施国家的信息基础设施建设,一场建设信息高速公路的热潮在世界范围内展开。

随着信息高速公路的建设,数字图书馆便应运而生,数字图书馆就是运行在信息高速公路上的“车”。有了信息高速公路和数

字图书馆,生活在地球上任何角落的人们,在任何一台终端上可以通过网络方便地访问遍布全球的数字图书馆,检索自己所需的信息数据,实现知识快速传输和信息资源的真正共享。

2004 年底 Google 宣布, 将把多家世界一流大学至少 1 500 万本藏书上网。也就是 Google 打算将 1 500 万多本图书数字化, 读者可以通过 Google 搜索引擎浏览图书馆的书本, 跟现在使用浏览器的方法一样, 只要输入关键字, 有关书本的链接就会显示出来, 方便查阅。其实 Google 的做法并不新鲜, 网络上早就有了不少的数字图书馆。但由于 Google 对图书数字化的规模很大, 所以它会产生深远的影响, 将改变人们的阅读习惯。只要通过网上搜寻, 就可以浏览远在哈佛、密歇根、牛津和斯坦福大学图书馆的藏书, 而不用亲自跑到图书馆去。斯坦福大学图书馆长凯勒 (Michael Keller) 说: “在 20 年内, 世界上的知识都会数字化, 所有人可在网上免费阅读图书馆的文献, 正如平时在图书馆免费借书一样。”然而, 图书数字化的意义, 绝不仅仅是把书本搬到了网络上, 它还将大大加快知识的传播速度, 改变人们获取知识、接受教育的方式。同时, 它还将改变人们的工作方式, 使跨国界的合作成为可能。

同样, 为迎接“网络经济”的到来, 中国政府也积极参与信息高速公路和信息数字化的建设。目前, 就公用数据通讯平台来说, 中国已基本建成了覆盖全国的多功能、多层次、高效先进、完整统一的公用数据通讯平台。相继建设了中国公用分组交换数据网 (CHINAPAC)、公用数字数据网 (CHINADDN) 和中国帧中继宽带多业务网 (CHINAFRN); 就 Internet 服务提供商来说, 国家级的 Internet 提供商中国就有 4 个: 中国公用计算机互联网络 (ChinaNct)、中国教育和科研计算机网 (CERNet)、中国科技网 (Cst-

Net)、中国金桥网(ChinaGBNet)。

覆盖全国的计算机通讯网络为我国数字图书馆的发展提供了良好的发展环境。中国高等教育数字图书馆建设项目应运而生。它旨在打破全国高校图书馆一校一馆独立发展的模式,将把全国高校的图书馆连接成一个整体,使读者可以实现每周7天、每天24小时通过网络,方便地查询到全国高校图书馆甚至国外一些图书馆藏书信息。

中国高等教育数字图书馆建设的目标是要汇聚全国乃至全球重要的教育与科研文献信息资源,整合各类数字化服务手段,同时依托教育科研网,支撑所有高校图书馆的文献信息服务,提高高校图书馆的整体服务能力和服务质量。目前,这一项目已显现出它的功效,读者正在利用网络,在数字图书馆中汲取知识。

1.2 数字图书馆的定义及特征

1.2.1 数字图书馆的定义

尽管数字图书馆已经成为图书馆学界的一个研究热点,然而,关于“数字图书馆”却至今还没有统一的定义,最典型的有以下几种观点。

研究图书馆协会 (Association of Research Libraries, ARL) 于1995年10月给出了一个要素列举式的数字图书馆定义:

“数字图书馆定义有多种,电子图书馆和虚拟图书馆常用作同义词。数字图书馆的公认要素有:数字图书馆不是一个单一的实体;数字图书馆需要链接许多信息资源的技术;多个数字图书馆惯用语信息机构之间的链接对最终用户透明;全球范围存取数

字图书馆与信息服务是一个目标;数字图书馆的收藏并不局限于文献的数字化替代品,还扩展到不能以印刷形式表示或传播的数字化人造品。”

1995年美国联邦信息与应用项目所用的定义是:“数字图书馆是向用户群体提供的便于查找、利用庞大的、经过组织的信息和知识存储的手段的系统。”

1997年,国际“分散式知识工作环境”会议指出:“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,它更是一个环境。它将收藏、服务和个人带到一起,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。”

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:“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,从而使这些数字式资源能够快速且经济地被特定的用户或群体所利用。”

大英数字图书馆认为:“数字图书馆是通过使用数字化技术获取、存储、保护和提供信息与提供信息与查询途径而被广泛认可的虚拟描述,而不论这些信息的当初形式如何。”

美国著名数字图书馆专家阿姆斯认为:“数字图书馆是具有服务功能的整理过的信息收藏,其中信息以数字化格式存储并可通过网络存取。该定义的关键在于信息是整理过的。”

中国科学院高文等人认为:“数字图书馆是以电子格式去存储海量的多媒体信息并能对这些信息资源进行高效的操作,如插入、删除、修改、检索、提供访问接口的信息保护等。”

上海图书馆刘炜等人认为可以给数字图书馆下一个比较宽泛的定义：“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。”

有关数字图书馆的定义还有：

数字图书馆是能将各种文字、图片、音频、视频等多媒体的文献信息内容，以数字化存储与管理，通过控制系统标识数字对象，通过网络浏览、查询、检索和传输各种信息，采用权限管理技术保护知识产权，通过 Web 发布数字化信息的网上图书馆。

数字图书馆是以计算机技术、网络通信技术、数字技术和文献处理技术为基础的大型数字化文献资源信息系统。

数字图书馆是一种电子图书馆，它是为人数量多而又处在不同地理位置的用户提供能够方便地利用大量的、分散在不同贮存处的电子物品的虚拟环境。这些电子物品就是指计算机能够读取的文本、图形、声频、视频等。

数字图书馆是一个大系统，它拥有分布的、大规模的和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，获得自己所需的最终信息。

数字图书馆建立在图书馆内部业务高度自动化基础之上，不仅能使本地和远程用户联机存取其馆藏联合目录以查询传统图书馆馆藏，而且也能使用户通过网络联机存取图书馆以外的其他信息资源的现代化图书馆。

关于数字图书馆不同组织机构、不同专家学者，从不同的角度出发，提出了很多定义，这里就不一一列举了。值得指出的是，数字图书馆与传统图书馆有着不可分割的联系。传统图书馆是数千年“前数字化时代”人类社会知识文化的结晶，数字图书馆虽然可以把传统图书馆中各种载体的文献信息内容数字化，但却不能

替代文献载体本身,也不能提供传统图书馆特有的阅览环境等。数字图书馆不仅要选择性地对已有的文献资源进行数字化,而且要处理新生的数字信息资源并开辟利用图书馆的新渠道、新方式和新技术。因此,没有必要摆脱传统图书馆去另建新的数字图书馆,传统图书馆是数字图书馆的基础,数字图书馆是传统图书馆的发展,二者相互结合,构成混合图书馆。

1.2.2 数字图书馆的特征

把握数字图书馆的特征有助于更全面准确地理解数字图书馆的定义,进而掌握图书馆的本质内容,从而更好地利用数字图书馆。

经过多年的发展,数字图书馆从逻辑上经历过以技术为主导、以资源为主导和以服务为主导的三个先后相继的发展阶段,数字图书馆也相应地出现了技术主导型、资源主导型和服务主导型三种模型。由此可见,数字图书馆的本质特征包括技术、资源、服务三方面的内容。

1.特色技术

由上述数字图书馆的定义可知,计算机技术和网络技术是数字图书馆建立、发展和利用的基础。数字图书馆除采用通用的计算机技术和网络技术以外,还有自己的特色技术,这是数字图书馆区别于其他技术领域的特征。现有特色技术包括分布式资源与运行管理技术、海量信息存储与组织技术、多媒体信息标引与检索技术等。目前,建立数字图书馆特色技术的相关标准是数字图书馆研究和建设的重要内容之一。

当前,国内外关于数字图书馆特色技术标准的研究和制定工

作发展迅速,形成了一系列相关的技术标准,对数字图书馆的发展起着重要的推动作用。在数字图书馆研究与建设方面,中国拥有与西方发达国家站在同一起跑线上的机会和不相上下的技术标准,这对于参与国际数字图书馆技术标准的制订极为有利。标准的重要性不言而喻,众所周知,全球共同遵循的 TCP/IP 协议,才有因特网的今天。同理,数字图书馆要向更高、更远的方向发展,也需要一套国际公认的特色技术标准。

2. 数字资源

数字资源是指图书馆中所有数字形式的信息资源,包括原先有用纸张形式存储的信息转换电子计算机中的数字化信息和本来就是以数字形式出版的信息。这些数字资源是数字图书馆的“物质”基础,也是数字图书馆有别于传统图书馆的一大特征。数字资源为压缩存储空间,改进信息的组织方式,提高检索速度,方便用户远程检索等奠定了基础。数字资源从不同的角度出发,具有不同的划分方式:

(1)从文献类型角度划分。数字资源包括期刊、图书、工具书、专利说明书、学位论文、视频资料、声频资料等多文献类型。

(2)从揭示文献内容的程度划分。虽然,数字图书馆的目的是直接提供读者所需的最终信息,而不是二次文献信息,但数字图书馆也需要有书目数据、索引文摘等二次文献信息,因此,二次文献信息也是数字图书馆的数字资源的一种类型。

(3)从文件格式划分。数字资源包括从位图形式的页面到经 SGML 编码的特殊文本文件,甚至 CD-ROM 中的信息或本地局域网中的资源等。

另外,能同时处理多媒体化的数字资源是数字图书馆在技术上的一个典型特征。