



苏东出 石晓东 孙 萍 ◎ 著

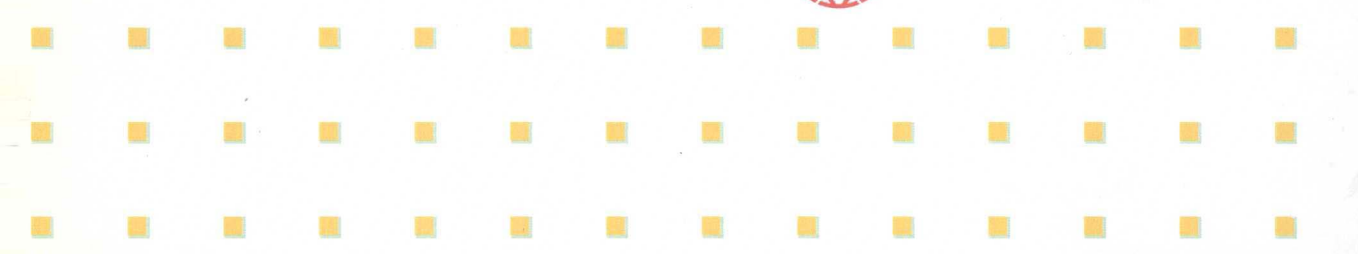
数字图书馆技术

导论

SHUZI TUSHUGUAN JISHU DAOLUN



西安地图出版社



数字图书馆技术导论

苏东出 石晓东 孙 萍 著

西安地图出版社

图书在版编目 (CIP) 数据

数字图书馆技术导论 / 苏东出, 石晓东, 孙萍著. —西安: 西安地图出版社, 2008. 3
ISBN 978-7-80748-244-4

I. 数… II. ①苏…②石…③孙… III. 数字图书馆—概论 IV. G250.76

中国版本图书馆CIP数据核字 (2008) 第 033296 号

数字图书馆技术导论

苏东出 石晓东 孙萍 著

西安地图出版社出版发行

(西安市友谊东路 334 号 邮政编码: 710054)

新华书店经销 陕西地质印刷厂印刷

787 毫米×1092 毫米·1/16 开本 13.5 印张 347 千字

2008 年 4 月第 1 版 2008 年 4 月第 1 次印刷

印数: 0001~1000

ISBN 978-7-80748-244-4

定价: 26.80 元

前 言

作为一种新兴的知识存储与交流方式，数字图书馆吸引了很多学者和专家的注意力。所谓数字图书馆，就是利用先进的数字化技术，将传统的图书馆馆藏文献资料以计算机能处理的数字信息的形态，通过某种逻辑结构存储、组织起来，并通过Internet等计算机网络，让人数众多又分布在不同地理位置的人们能方便地检索、利用的一项活动，是一种新型的知识存储与交流方式。由于数字图书馆以网络和高性能的计算机为依托，向人们提供比传统图书馆更先进、更广泛、更方便的服务，从根本上改变了人们获取信息、组织信息、利用信息的方式，作为一种新兴的网络知识服务方式，数字图书馆在整合传统资源服务于大众方面具有无法比拟的优势，必将在建设信息强国的道路上发挥主导性的作用，已受到了世界各国的重视。

本书首先讲述了数字图书馆发展的现状以及在当今知识经济时代的特殊意义，介绍网络数据库以及开发接口技术、目前流行的Web开发技术ASP.NET、数字图像处理技术。结合这些技术给出一个简单的数字图书馆系统实例，详细并有重点地讨论了前几章介绍的技术在数字图书馆系统中的具体应用和实现。其次探讨了数字图书馆数据安全、著作权、人才培养问题，力求使读者了解数字图书馆在实施过程中所面临的疑难，并对所存在问题的解决有所思考。最后介绍了新理论新技术在数字图书馆领域的应用，它们给数字图书馆带来了新的发展方向，促使其更好地发展，也为用户带来更方便的服务。

本书第一～四章由石晓东执笔，第五～八章由苏东出执笔，第九～十二章由孙萍执笔。

本书实现的基于B/S和C/S架构的数字图书馆系统DlibSys，是对数字图书馆系统的尝试。当然，限于硬件环境、时间、作者的研究水平等因素的制约，所做的工作还比较简略，有待于在今后的工作中进一步完善。

编者

2007年10月

目 录

第一章 数字图书馆的基本概念	1
第一节 数字图书馆的产生.....	1
第二节 数字图书馆的有关定义.....	4
第三节 数字图书馆的特征.....	9
第四节 数字图书馆的功能.....	12
第五节 数字图书馆带来的变革.....	13
第二章 国外数字图书馆的发展概况	20
第一节 美国数字图书馆的发展状况.....	20
第二节 英国的数字图书馆试验计划.....	23
第三节 日本数字图书馆项目研究状况.....	23
第四节 全球数字图书馆联盟计划.....	26
第五节 其他国家的数字图书馆项目.....	28
第三章 我国数字图书馆的建设	31
第一节 我国数字图书馆研究现状.....	31
第二节 数字图书馆建设中需解决的问题.....	40
第三节 我国数字图书馆建设的必要性.....	43
第四章 数字图书馆的体系结构	46
第一节 数字图书馆系统的信息结构.....	46
第二节 数字图书馆信息标准.....	50
第三节 数字图书馆系统的结构模型.....	55
第四节 数字图书馆系统涉及的技术.....	58
第五章 Web 数据库	64
第一节 Web 数据库技术	64
第二节 Web 数据库开发与集成的关键技术	67
第三节 Web Service 技术	71
第六章 系统开发平台	78
第一节 dotNET 技术简介	78
第二节 ASP .NET 技术简介	86
第三节 ADO .NET 技术简介	90

第七章	数字图像处理 DIP 技术	95
第一节	DIP 技术概述	95
第二节	图像的运算	101
第三节	图像增强与压缩编码	106
第八章	数字图书馆的一个应用实例——DlibSys 系统的实现	112
第一节	系统概述	112
第二节	数据存储的实现	113
第三节	数据处理模块的实现	119
第四节	数据发布模块的实现	132
第五节	Dlib Reader 的实现	137
第六节	移动服务模块的实现	141
第七节	目前存在的问题以及后续开发计划	146
第九章	数字图书馆数据安全研究	147
第一节	数据安全简介	147
第二节	软件技术预防机制	149
第三节	硬件技术预防机制	152
第四节	人为因素预防机制	155
第五节	数字信息资源的长期保存机制	157
第十章	数字图书馆的著作权研究	162
第一节	网络环境下著作权保护	162
第二节	数字图书馆中的著作权保护问题	165
第三节	数字图书馆著作权的解决方案	171
第十一章	数字图书馆的人才培养	175
第一节	数字图书馆人才管理	175
第二节	数字图书馆人才管理策略	178
第三节	学位教育培养	184
第十二章	数字图书馆的发展趋势	190
第一节	知识管理对数字图书馆的影响	190
第二节	先进技术对数字图书馆的影响	192
第三节	数字图书馆服务趋势	202
参考文献		208

第一章 数字图书馆的基本概念

第一节 数字图书馆的产生

从1807年德国的M.施雷廷格提出“图书馆学”这个专门名词,到1887年美国的M.杜威创建的哥伦比亚大学图书馆管理学院的诞生,标志着图书馆学作为一门近代科学的正式确立。100多年来,人类图书馆在经历了以纸质印刷图书资料为主的传统图书馆、以建立“机读目录数据库”并辅以计算机管理全面提高业务工效的自动化图书馆时代之后,随着当代四种先进技术即计算机技术、通信和网络技术、高密度存储技术以及多媒体技术的飞速发展,人类正在迈入旨在建立有序的信息空间,实现真正意义上的信息资源共享为目标的新领域——数字图书馆,它是21世纪图书馆发展的主要方向。

从宏观角度来看,图书馆事业的发展和演变,都离不开内部环境的驱动力和社会环境的推动力。内部环境主要是指社会对图书馆的需求产生的驱动力,促使其改变传统图书馆的服务环境,实现图书馆的自动化。社会环境的变化主要包括社会经济结构、信息技术结构、文化结构的变动,进而推动了图书馆的发展和演化,促成了数字图书馆的出现。

一、图书馆内部环境的变革

1969年,美国国会图书馆正式发行MARC II型的机读目录,使图书馆走向了标准化、规范化生产机读目录的新时期。随之出现了图书馆自动化集成系统,代替了传统图书馆的手工作业和业务流程,改善了图书馆的读者服务手段,提高了服务效果和工作效率。70年代之后,图书馆界、情报界又出现了一大批联机编目中心和大型文摘与检索系统,如OCLC、WLN、RLIN、DIALOG、ORBIT、MEDLINE、BRS等,这些系统较深层次揭示了馆藏图书、期刊的内容,生成了大规模的书目数据库。这时,图书馆已进入书目数字化的阶段,也标志着图书馆有能力推动图书馆自动化系统的发展。进入80年代以后,特别是局域网络技术迅猛发展与成熟,图书馆建立了以局域网络为基础的网络系统,通过局域网络与广域网络连接,实现了远程查询图书馆的机读目录和大型文献信息数据库,这一时期为文献数字化的网络查询、检索、传输打下了良好的基础。90年代以来,世界先进国家的图书馆已经完成了图书馆自动化书目时代的任务,正加紧数字图书馆的建设,进行大容量、多媒体信息处理平台建设和文献信息资源的数字化转换工作。如美国国会图书馆,1995年就开始进行国家数字图书馆项目建设,现已完成500万件文献图片的数字化转换处理。我国国家图书馆现已完成了4 700万页文献的扫描和制作工作,并在网上提供服务。上海图书馆在几年前就开始对古籍文献进行扫描加工,建立了古籍电子数据库。这一时期,数字图书馆的试验、研制取得了许多重大成果,如快速扫描与压缩技术,图像、声音的采集存储与检索技术,文献的标识与识别技术,版权保护技术等等。图书馆经过二三十年的努力,采用信息技术,实现了卡片目录、书目索引、文摘的数字化,正进入对图书文献内容的全文数字化转换阶段,逐渐改变了图书馆的内部环境,适应了社会对图书馆的需求。

二、社会环境的变革

1. 社会经济结构的变动

在现代经济结构中,信息成为资本,信息资源成为了战略资源,产生了新的经济理论

——知识经济。另一方面,由于信息技术的迅速发展,产生了新兴的信息产业,而且快速成为高新技术的主导产业,其中信息有序化处理、信息服务业已成为信息产业中一个重要的分支,为图书馆的发展提供了一个很好的发展机遇,加快了图书馆信息服务自动化、网络化的进程。

2. 技术结构的变革

在当代信息技术结构中,计算机、通信、网络、数字技术已成为当代社会变革的最大的技术支柱,这些技术的成熟与发展,也为图书馆的自动化、网络化、数字化的研制、试验、运行提供了技术基础。在这些技术的支持下,图书馆建成了一系列高质量的文献信息服务系统、大型的数据库系统以及网络信息系统。电子出版物光盘的出现和应用,为电子图书馆的产生提供了条件;大容量存储设备及网络的应用,为虚拟图书馆提供了一个巨大的信息资源宝库;因特网的应用与普及,为数字图书馆的产生提供了实实在在的网络环境,为图书馆的信息资源开发与利用提供了有利条件。总之,信息技术将促进图书馆在更广泛的时空中,向网络化、数字化方向发展。

3. 文化结构的变化

在当代文化结构中,逐渐形成了文化、教育、科技三位一体的文化,称为大文化。信息技术改变了文化艺术的表现形式,出现了以数字冠名的设备、影视、多维活动图像、音乐等。如数字照相机、数字摄像机、数字电视、数字影视片、数字音像、数字报刊等,这些都属于电子信息文化的组成部分。这种信息服务必然是数字图书馆才能做到的,而且也是现代图书馆信息服务业的信息增值和经济增长点之一。

4. 信息资源已开始全面电子化、数字化

信息内容已开始全面数字化,随着越来越多的数据库和信息系统的广泛应用,文献资料大规模地上网。就我国公共图书馆而言,近几年来,国家图书馆、上海图书馆、辽宁图书馆、中山图书馆、深圳图书馆、南京图书馆、桂林图书馆等许多图书馆都在用不同的方法开展文献资料数字化工作。清华大学出版的中国学术期刊(光盘版)已收录国内重要学术期刊 26 000 多种,论文约 40 万篇(一年),我国的《人民日报》等 20 多种重要的报纸都可在网上查询。所以说,因特网是一座巨大的数字图书馆,也是一座全球的虚拟图书馆。

三、数字图书馆的出现

尽管数字图书馆的概念萌芽可以追溯到半个世纪以前,即世界上第一台电子管计算机刚刚诞生时,就有对“数字化信息”及其计算结果的一些提法,但真正出现是在近十年以来。在新技术的带动下,当图书馆的内部环境和社会环境产生巨大变革后,一系列新生事物,如因特网、多媒体、万维网、国家信息基础设施等产生之时,“数字图书馆”也迅速成为一个时代的特征之一。“数字图书馆”这一名词的出现与美国政府提出兴建国家信息基础设施和因特网的迅速普及处在同一时期,显示了它们之间不可分割的联系。

数字图书馆作为现代信息技术环境下应运而生的新型图书馆,目前还没有一个明确、完整的定义。数字图书馆中的“数字”,是由英文“Digital”翻译而来。一般而言就是利用先进的数字化技术,将传统的图书馆馆藏文献资料以计算机能处理的数字信息的形态,通过某种逻辑结构存储、组织起来,并通过Internet等计算机网络,让人数众多又分布在不同地理位置的人们能方便地检索、利用的一项活动。由于数字图书馆以网络和高性能的计算机为依托,向人们提供比传统图书馆更先进、更广泛、更方便的服务,从根本上改变

了人们获取信息、组织信息、利用信息的方式,因此受到世界各国的重视。

随着现代信息技术在图书馆的广泛应用,近年来,有人总结了国内外图书馆信息技术的应用与发展情况,提出了图书馆发展的三阶段理论。

1. 图书馆自动化发展的初级阶段

大约从 20 世纪 60 年代末、70 年代初开始,以美国国会图书馆正式发行 LCMARC 机读目录为标志,它在北美得到广泛的应用,开创了书刊机读目录在世界上正式使用的新时期,使图书馆正式步入了自动化的阶段。1972 年国际图联推出了国际机读目录格式 UNIMARC,随之,各国也相应推出了本国的机读目录格式。如英国的 UKMARC、日本的 JMARC、中国的 CNMARC 等等。

这一阶段,图书馆计算机管理集成系统开始投入使用,其基本模块是图书采购、图书编目、图书流通、期刊管理、书目查询等。它已初步实现了让读者利用计算机查找所需资料 and 实现图书业务自动化管理的过程,在一定程度上解放了图书馆工作的一部分劳动力,使图书馆管理更加科学化,从而提高了对读者的服务质量。

在这一阶段,国外已出现了一些图书馆计算机协作网络,如联机编目网络等,另外还出现了一批文摘社和信息中心,这些系统为对文献进行较深一步的加工作出了贡献,让读者可以用联机方式得到许多会议录和杂志中的文章题名和摘要等,已为第二阶段的电子文献信息服务奠定了基础。

2. 网上电子文献信息服务的新阶段

20 世纪 80 年代中期,随着计算机网络技术的发展和广泛投入应用,发达国家的图书馆都开始使用局域网络向读者提供网上信息服务。80 年代末,光盘数据库开始投入应用。90 年代初,因特网开始投入商业运行,这些都为图书馆信息技术的应用提供了新的条件。于是,图书馆很快步入一个全球性、整体性网上电子文献信息服务的新阶段。在这一阶段的主要标志是:

(1) 局域网络在图书馆的广泛应用

图书馆不但建立起馆内局域网,还连接到校园网、城域网、全国性网络,开始利用 LAN 进行网上联机编目、网上图书采购,为读者提供联机目录查询、网上馆际互借、传送电子出版物的索引及摘要以至全文。

(2) 光盘数据库在图书馆的广泛应用

1984 年,荷兰飞利浦公司和日本索尼公司在 COMDEX 年会上首次展出他们的 CD-ROM 驱动器和光盘数据库,随后美国图书馆公司推出第一个商用产品——BIBLIOFILE,存有美国国会图书馆的书刊机读目录数据。近十年来,CD-ROM 光盘及数据库产品已广泛地在图书馆投入使用。据不完全统计,到 2002 年为止,全世界已有数十万种 CD-ROM 文献光盘数据库,有各种参考工具书、辞典、百科全书、指南、索引、摘要、全文数据库和多媒体数据库。包括自然科学、社会科学、工程技术和经济管理等各个方面。当光盘由单机使用发展到光盘网络,它的作用就更为明显,读者可以在图书馆、办公室、实验室甚至在家里查找图书馆的光盘数据库。

(3) 图书馆自动化集成系统向网上电子文献信息服务的方向发展

70 年代的图书馆自动化系统,大都是单功能系统,80 年代中期,逐渐向集成系统过渡;在数据结构方面,由文件系统转向数据库管理系统;在系统结构方面,集中式的数据库转

向分布式的网络结构,形成了客户机/服务器的访问方式;在功能方面,除了一般馆内业务处理模块外,还有与外界数据接口的功能,让读者很方便地利用系统得到最新杂志目次内容的服务,甚至可得到所需文章的全文传递服务;连全世界最大的联机目录中心OCLC也推出了网上电子文献服务。由此可见,网上电子文献信息服务在90年代已成为图书馆和信息中心不可抗拒的潮流。

(4) 因特网的迅猛发展,把图书馆电子文献信息服务推向全球化

因特网不仅是现阶段国际信息高速公路的雏形,而且是目前世界上最大的信息宝库,包含有图书馆、信息中心的大量文献信息资源,人们可以像过去利用传统图书馆一样,通过联机访问各种电子信息资源。

目前,通过因特网可以查找到国内外几千个图书馆的书目信息,可访问世界上许多著名的情报数据库,如DIALOG、BRS、MEDLINE等;可访问许多免费的信息数据库,也可访问世界上20多万个万维网服务器上各单位的主页,还可以获得大量的全文资料和多媒体的资料。因特网的发展,已把图书馆电子文献信息服务推向全球化。

3. 图书馆发展的高级阶段——数字图书馆阶段

近年来,由于因特网的广泛应用,传统图书馆的观念开始向电子化、网络化和虚拟图书馆的概念转化,一大批电子出版物和全文数据库被存储在图书馆和各类信息中心,致使图书馆的形态发生重大改变,于是人们提出“数字图书馆”的新概念,即将图书馆文献信息以数字化的形式表现和获得。这将是图书馆发展的高级阶段。

目前,国内外许多机构和研究部门,都在从事数字图书馆项目的研究开发工作。有一些数字图书馆的开发技术平台已基本成熟并投入市场,一些试验型数字图书馆系统也研制出来,正准备投入使用,许多单位都在开展数字资源库的建设和前期准备工作,在一定范围内开展的“网上读书”活动已受到许多读者的青睐。

这三个阶段并不是截然分割开的,它们相互交错,相互渗透,前一阶段又为后一阶段做了一些技术准备。随着需求和技术的发展,从低级阶段向高级阶段发展。如第一阶段出现的机读目录,实质上是图书馆馆藏目录的数字化,也是第二阶段电子文献信息服务的重要组成部分。在第二阶段中,已对数字图书馆的技术做了大量研究,出现了一些数字图书馆雏形,如在因特网上的许多WWW服务器,已可以提供文本、图形、声音、动态图像和多媒体资料查询等。而且,对各种文献载体的数字化,大容量数据的存储和管理技术,各种文献的访问及传输技术等均为第三阶段的数字图书馆技术走向实用化做了准备。

第二节 数字图书馆的有关定义

对信息技术在图书馆的应用,产生了不少新名词,如电子图书馆、虚拟图书馆、网络图书馆、无墙图书馆、多媒体图书馆、环球图书馆和数字图书馆等称谓。数字图书馆这一新颖的论题得到政府部门、信息技术界人士、信息资源部门的关注和支持,使得众多报刊纷纷登文讨论,美国专门出版了数字图书馆杂志,并在美国的奥斯汀举办了第一届数字图书馆国际会议。1996年,美国计算机协会、信息科学学会等联合举办了“首届ACM数字图书馆国际会议”。会议讨论了数字图书馆的范围、定义、特点和功用、信息结构,电子出版物的控制与管理、知识产权保护、WWW与导航技术以及与数字技术相关的技术挑战、因特网的发展趋势等等课题,这些课题在广义上都涉及了数字图书馆的范畴与内容。目前,由于数

字图书馆领域涉及多种学科、多种技术,而且是一个巨大的系统工程,所以很难界定其含义,也没有一个明确的、完整的、规范的定义。下面就这些概念作一介绍。

一、电子图书馆

首先,我们从概念所依据的技术基础谈起,任何一种概念的提出都离不开当时所处的社会环境和技术基础。电子图书馆的概念最早是由美国科罗拉多州的一个区图书馆馆长Dowlin提出来的,他在1984年出版的《电子图书馆:前景与进程》一书中写到:“所谓电子图书馆是一个最大可能提供存取信息的并使用电子技术对信息资源进行管理的机构。”

Dowlin所描述的电子图书馆主要还是图书馆自动化的内容,即书目的生成、联机检索等。其可贵之处就是利用邮政线路,实现图书馆间的联机存取,其技术基础仍然是书目型数据库系统、联机查询技术、邮电网络技术、局域网络技术。当时,这些技术在美国都是比较成熟的技术,所以这一时期的电子图书馆并不是真正意义上的电子图书馆。

到了20世纪90年代初,大量的电子出版物等新型的电子信息资源涌现,1990—1995年统计,CD-ROM出版物总种数为30 917种,是当时电子出版物的主流产品。这些光盘可存储文字、声音、动画等多种形式的信息,而且以全文为主体,发行了大量的电子图书、电子期刊、电子报纸,同时,可通过网络对因特网进行查询,再现原文、图像等。这时的技术主要是扫描技术、压缩技术和还原技术,但多媒体技术仍处在试验和初级运行阶段。

从1993年开始,电子图书馆的概念逐渐被数字图书馆所代替。那么,电子图书馆到底应如何定义?我们还是从技术角度来定义:电子图书馆是利用现代电子技术、通信技术、高密度存储技术等新技术和手段建成的一种大型的交互式信息系统,它既可组织、存储各种载体的书目型或全文型的文献信息,又可使用户通过局域网络或互联网络查询和获得各种载体的文献资料。也就是说电子图书馆是一种概念,一种扩大了功能的、在网络环境下运行的一个以电子出版物为主体的图书馆自动化系统。

二、虚拟图书馆

对虚拟图书馆,国外许多学者还有不同的认识和不同的定义,具有代表性的有以下几个。

Kaye认为:“虚拟图书馆指的是一种环境,其各组成部分协同作用,提供智能化的、实实在在的信息存取途径,其价值在于它完全从每一个用户的独特观点出发,构建系统框架以满足用户的信息需求。”

Poutler认为:“虚拟图书馆不一定非得基于一个实际的图书馆,它可以是存储在通过网络连在一起的许多图书馆中的资源和服务的混合物。”

Saunders在一篇论文中提到:“虚拟图书馆是一个系统,通过使用本地图书馆的联机目录,一所大学或网络计算机的网点,用户可以透明地连接远程的图书馆和数据库。”

Beiser观点认为:“虚拟图书馆即没有围墙的图书馆。它以电子方式将世界范围内的图书馆、个人、机构及商业公司连接起来,并提供检索其存储的学术信息资源。这些资源不仅包括正式的图书馆信息,还包括数据库、电子文本、多媒体对象,以及数以百万计的相互作用的人类智力因素。”

上述观点都强调通过网络查询和传输信息,可获得各种各样的信息资源,是一种没有围墙的图书馆。用户不必关心信息放在何处,好像他自己拥有一座信息量无比巨大的图书馆,这就是用户“虚拟的感觉”。“虚拟”一词,据《现代汉语词典》的解释,是“假设的”、

“虚构的”意思。因此虚拟图书馆就应该是一种假设的、不以一种物理实体存在的虚构的图书馆。但从信息资源的意义上讲,它又是一种实实在在的,可通过网络实现资源存取的大规模信息资源系统。对读者来说,实际上它是通过自己的网页连接世界上可连接的图书馆网页而形成的“虚拟图书馆”。

由此可见,虚拟图书馆是在计算机网络上对分布于各地的各种信息资源进行动态搜寻连接的一种行为方式,而不是一种图书馆实体形态;它的出现,并没有否定传统图书馆的继续存在。

虚拟图书馆虽然不是一个物理概念,不是一个固定的处所,也不是一个独立的实体,但它仍然需要一些基本的物质条件、技术方法和非物质因素才能构成。组成虚拟图书馆的要素包括:数字化信息及其载体、计算机及其相关设备、通信网络或信息通道、网络用户、信息技术、网络协议和网络管理等。缺少任何一个要素,虚拟图书馆都不能存在或不能运行。

虚拟图书馆与传统图书馆之间的关系,不是替代的关系,而是互相依赖、互相促进的关系,如果没有传统图书馆选择、收集、加工文献信息,虚拟图书馆中的信息资源就会匮乏。反之,如果没有虚拟图书馆提供新的信息环境,传统图书馆也不可能突破原有工作的局限,有限的馆藏和服务就难以充分满足用户的需求。可见,虚拟图书馆是建立在传统图书馆基础之上的。同时,网络化的虚拟图书馆也为传统的图书馆提供了进一步发展的机遇。

三、网络图书馆

随着因特网在全球的普及,一种新的通过计算机网络获取信息的资料库——“网络图书馆”正在逐步形成。所谓网络图书馆,并非一种实体的图书馆,而是“利用电子网络远程获取信息与知识的一种方式”。在这种方式下,信息存储和检索利用的地理界限已经被打破,所有人都能通过计算机网络随意查询分布于世界各地的数据、图表、文献等各种信息。由于越来越多的数据库和信息系统不断加入网络,网上信息的种类和数量不断增加。因特网逐渐成了目前世界上资料最多、门类最全、规模最大的信息资料库,因而有人称其为“全球最大的图书馆”。

显然,这种以网络为基础的所谓“网络图书馆”,与传统意义上的图书馆在形态上有本质的区别:它没有馆舍、没有藏书,人们也不必亲自到图书馆去,只要在办公室或自己的家中打开与网络相连的计算机终端,即可获取信息。由于这种网络图书馆能够提供庞大的信息资源和方便快捷的服务,因而对传统图书馆工作提出了严峻的挑战。一些学者甚至预言,在不久的将来,网络图书馆将取代传统图书馆,认为这是“今后图书馆发展的归宿”。

借助于计算机网络,一个图书馆的信息资源可以向其他图书馆和用户提供服务,同时该馆也可以接受其他电子图书馆提供的信息服务。只有这种在网络的支持下,在不受时空的限制远程利用他馆的信息资源和服务的情况下,才构成了一种不是实体的图书馆却胜似实体图书馆的网络图书馆环境。所以,与其说网络图书馆是一种图书馆形态,不如说它是一种信息环境或信息空间。在这种环境下,网络终端无论延伸到何处,其终端用户都可以自由地使用网络上的各种信息资源,而不管这些资源来自于何处。显然,这种网络图书馆环境的确给我们带来了利用全球信息资源的便利,它向我们提供了一种将各个图书馆和信息服务机构联系起来,在更大范围内获取文献信息的手段。

四、数字图书馆

数字图书馆是一个正在成长的新生事物,国内外对其概念的界定还没有形成一致的意见,相关的描述多达数十种。本书无意于评判这些定义,但从历史的角度作简要回顾,以便帮助大家“数字图书馆”概念的进一步理解。

按福克斯的说法:“数字图书馆是一种基于纸质的图书馆外观和感觉的图书馆,但这类图书馆的资料都已被数字化并存储起来,而且能在网络环境中被本地和远程用户存取,还能通过复杂和一体化的自动控制系统为用户提供先进的、自动化的电子服务。”从福克斯的观点来看,数字图书馆仍然是传统图书馆的继续,其不同的是文献被数字化,并可通过网络存取,还能通过一体化控制达到管理和规范。

早期研究引用较多的一个定义来自美国研究图书馆协会:“数字图书馆不是一个单独的实体,需要有关技术提供到其他资源的链接,该链接对用户应该是透明的,目标是做到任意检索,数字馆藏应超越传统馆藏而不能成为其替代品。”

有学者认为:“数字图书馆是一个数字化信息系统。它将把分散于不同载体、不同地理位置的信息资源以数字化的形式存储,以网络化的方式互相连接,提供及时利用,实现资源共享。”这种观点界定了数字图书馆是一个数字化信息系统。

也有人认为:“数字图书馆是计算机数据库或数字化对象的集合。”这个观点从宏观上界定了数字图书馆是一个数据库系统。

另一种观点认为:“所谓数字图书馆是这样一种图书馆,其全部或大部分馆藏都是以计算机可处理的形式存在的,以此来补充并逐渐代替构成当前图书馆馆藏主体的传统印刷本文献或缩微资料。”这种观点只局限于把传统图书馆的馆藏进行数字化的转换。

还有人提出:“数字图书馆是传统图书馆的扩展,既具有传统图书馆收藏、保护、提供服务功能,又提供集成化的、深层次的、数字媒体的远程数字图书馆服务。”这种观点认为数字图书馆是传统图书馆功能的补充,只是馆藏范围、服务方式上的深化和网络化。

1995年5月在美国弗吉尼亚州,由美国政府的信息基础设施技术与应用组主办的“数字式图书馆研讨会”上,由斯坦福大学计算机系统实验室主任Hector Garcia-Molina教授及加州大学伯克利分校校长办公室自动化主任、当时的美国信息科学协会主席Clifford Lynch教授提交的一份供会议讨论的定向报告中的定义是:“数字图书馆为国家信息基础设施提供关键性信息管理技术,同时提供其主要的信息库和资源库。换句话说,数字图书馆是国家信息基础设施的核心。”按照美国信息基础设施的报告,将用以下三方面建设支持先进的应用:

- (1) 数千个信息资源库;
- (2) 宽带数据网和信息设施;
- (3) 先进的通信和信息存取服务。

Hector指出,这里的(1)就是指的数字图书馆,而不是其他大量已建成的一般信息库。由此可见,数字图书馆中的“图书馆”是一种广义词,实际上“数字图书馆”中包含的内容将远远超出传统图书馆涵盖的范围,数字化后的各类信息,都将是数字图书馆中宝贵的资源。由此可知,数字图书馆建设不是指某一个馆的自动化系统水平或某一个馆的馆藏资料数字化程度,而是一项技术极其复杂,规模十分庞大、数据资源极为丰富,需花费大量人力、物力、财力和相当长时间才能实现的一项浩瀚的知识工程。

通俗地说,数字图书馆就是虚拟的、没有围墙的图书馆,是基于网络环境下共建共享的可扩展的知识网络系统,是超大规模的、分布式的、便于使用的、没有时空限制的、可以实现跨库无缝连接与智能检索的知识中心。数字图书馆既是完整的知识定位系统,又是面向未来互联网发展的信息管理模式,可以广泛地应用于社会文化、终身教育、大众媒介、商业咨询、电子商务等一切社会组织的公众信息传播。

归纳上述数字图书馆的定义,虽然用词不同,表述各异,但都有共同点:馆藏文献的数字化存储和通过网络查询与传输;数字图书馆是传统图书馆功能的扩展,是数字对象的集合,而且具备了自动控制的功能。为了比较清晰地描述数字图书馆的概念与定义,我们不妨从技术角度和图书馆角度去解释数字图书馆的概念与定义。

从技术角度来定义: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。

数字图书馆概念的技术基础:包含图书馆自动化集成系统的接口与连接技术、数字获取技术、数字描述技术、数字对象标识与对象库技术、关联索引与检索技术、用户界面设计技术、版权管理技术和文献处理技术。

从图书馆角度来定义:数字图书馆是传统图书馆功能的扩展,它既具有传统图书馆的自动化系统功能,又是能使图书馆用户联机存取本地和远程数字化资源的现代化图书馆。所以,数字图书馆的概念包含:

- (1) 数字图书馆仍然是一个能感觉到的图书馆,必然具有图书馆的文献收集、加工、整理、保存、服务的基本功能。
- (2) 数字图书馆以计算机可处理的数字形式存储文献信息,而不同于传统图书馆以纸质为主的文献收藏方式。
- (3) 数字图书馆的数字化信息收藏范围从广泛性和深层性上远远超出传统图书馆,所以它是以文献资源内容为基础的系统。
- (4) 数字图书馆提供更广泛、迅速、便利、多种形式的服务。依托因特网,利用先进的数字处理技术和网络工作站为全球用户提供远程服务。
- (5) 数字图书馆是传统图书馆的补充,尽管有人预言数字图书馆将来可以代替传统图书馆,但实际上数字图书馆只是在馆藏范围、服务方式上的补充与深化,完全代替是不可能的。

数字图书馆的概念就是图书馆利用数字技术,将其馆藏文献数字化并提供有效的服务和管理,同时利用该技术,组织网上所有可用的数字化信息的连接和导航。这样,用户可以获得图书馆经过数字化的所有文献信息,包括联机采购、编目、公共查询;组织用户通过网络检索本馆的信息资源、访问外界数字图书馆和文献信息数据库系统,如电子期刊、电子图书、声像资料、动画片、影视片和多媒体资料等;连接图书馆自动化系统,继续管理图书馆的采、编、流、期刊等的服务工作;利用数字图书馆技术建立数字图书馆的网络环境和数字图书馆平台,连接到机关、厂矿企业、学校、办公室、家庭等单位 and 团体,让人们能方便地共享图书馆及其他资源;利用数字图书馆技术,办好网上图书馆、网上专题

服务、网上公告；利用数字图书馆技术，管理和保护知识产权。

数字图书馆有三个基本要素：数字化资源、网络化存取和分布式管理。

数字化资源：大量的数字化资源是数字图书馆的“物质”基础。数字图书馆中的数字化资源应该能够直接提供读者所需的信息，而不只是二次文献。然而也应该包括大量的查找信息和帮助信息，因为二次文献也可能是某些读者的最终需求，所以书目数据，索引文摘等应该是数字图书馆的有机组成部分。同时由于数字化把各种不同载体的信息统一于 0 和 1 两个简单的数字，图书、期刊、录音录像带、缩微品、光盘或者古籍、善本、手稿、碑帖、字画、X光片等等，在数字图书馆中都消失了原本的物理形态，多媒体自然也成了数字图书馆的特征之一。

网络化存取：高速的数字通信网络是数字图书馆的存在基础。数字图书馆依附于网络而存在，其对内的业务组织和对外的服务都是以网络为工具或载体，使得它得益于网络也受制于网络。只有利用网络所提供的一切便利和优势，才能最大限度地发挥数字图书馆的作用。

分布式管理：分布式管理是数字图书馆发展的高级阶段，它意味着全球数字图书馆遵循统一的访问协议，即全球数字图书馆将像现在的Internet连接网站一样，可把全球的数字化资源连为一体，组成一个巨大的图书馆，使组织和共享人类所有知识成为一种可能。目前关于数字图书馆最关键的技术研究和开发重点就在于此，这也是数字图书馆之所以聚集众多的人力、财力，受到各国特别是发达国家高度重视的最主要原因。

因此，数字图书馆最根本的着眼点应放在数字化资源建设、网络化资源存取和分布式资源管理上，逐渐实现全国乃至全球网上资源共享。

第三节 数字图书馆的特征

一、文献内容数字化

1. 馆藏文献内容数字化

采用数字技术，将传统图书馆的馆藏，包括印刷型文献、缩微制品和视听资料的内容逐步实现数字转换和处理，并存储在大容量、高密度的存储装置中，采用数字图书馆有关数据存储和标识技术，对数字文本、图像、视频、音频图书分级存储，用调度系统把它们有机地集成在一起。采用这种数字化存储方式。其信息存储量大，体积小，能充分节省存储空间。

2. 电子出版物的加工处理

电子出版物包括电子图书、电子杂志、电子报纸等。这些出版物是电子排版文本格式，对这一类已数字化的出版物进行格式转换，并对数据元对象和内容、章节、目次进行标识，以及对内容进行分类、主题标引，用超文本技术，把它们与正文连接起来。经这样处理后，这些电子出版物具有了检索功能。

3. 网上下载的数字化全文处理

网上下载的数字化全文，如果采用PDF文件，就不必对数据进行转换，如果全文采用纯文本格式，则要对其进行格式转换。这样，才能与馆藏文献数字化格式一致，并且有浏览、查询和检索功能。

4. 文献数字化中应注意的问题

文献内容数字化处理,包括馆藏文献、外部电子信息资源,网上资源和非网上资源的数字化和技术处理,要有目的、有计划地根据用户需求、馆藏的基础、馆藏重点发展计划、馆藏特色和地区或全国合作计划等因素来确定。国内外数字图书馆的试验项目告诉我们,在组织馆藏文献信息数字化时要注意以下几个问题:

(1) 选题要有针对性,具有特色。

(2) 选题的文献信息类型馆藏较齐全。

(3) 以一次文献信息加工为主,辅以二次书目信息的建立。

(4) 选题要简单,范围不要太宽,否则项目过大,短期内不能完成,势必影响今后的发展。如美国国会图书馆数字化起步就是选择对美国历史与文化研究有价值的文献信息,建成了《美利坚记忆》。

(5) 要注意保护作者的著作权。

二、利用计算机技术管理资源

自从计算机技术在图书馆应用以来,实现了对各类文献信息的加工、采集、存储、检索、传递和业务管理,推动了图书馆自动化的发展。进入数字图书馆阶段,我们要利用计算机来管理多媒体文献信息资源,特别是对视频图像和音频信息的数字化存储、加工、处理、图像检索、音乐选播和多媒体信息传输,从而使多媒体文献信息增值。管理的内容包括多媒体信息数字化,数字化信息的标识与描述,组织规范性加工与存储,存取服务的管理,知识产权、存取权限、数据安全管理等。最终利用计算机技术、通信技术、网络技术、高密度存储技术建成具有管理功能的数字图书馆。

三、信息海量化

数字图书馆是丰富的文化遗产和当代迅速发展的各种现代科学技术信息资源的重要基地。所以,一个数字图书馆的信息规模必须达到相当大的程度,才能体现数字图书馆的价值,尤其是信息增值。如美国国会图书馆的数字图书馆计划要求使数字化的信息规模达到100TB,相当于500万件文献。所以说,它是一座巨大的信息资源宝库。资源库可包括本馆、本地区、全国各类型图书馆、信息系统的资源,也包括因特网各类型信息系统的资源。

四、信息标准化

数字图书馆的建设是由全国范围内很多图书、情报、档案机构以及各种信息中心和文化设施等众多部门和单位共同参与的。它所管理的信息和知识包括了所有学科,信息类型繁多,信息组织复杂,要实现网络的互联互通,资源的共建共享,管理的井然有序,必须使用统一的标准、统一的格式组织信息。信息的标准化和规范化是实现数字图书馆资源共享的前提和根本保障。目前世界各国都在加紧制订相关技术标准以取得信息控制权。美国国家科学家基金会(NSF)在开始实施其数字图书馆研究计划时就规定了三个关键技术目标,其一就是网络协议和标准研究。围绕着GML/XML/HTML等应用环境已产生多种元数据规范,这些规范使数字图书馆的信息储备、描述与标识检索查询、交换和使用走向标准化。标准化之所以是数字图书馆的基本特征在于它强调统一标准协议的重要性。只有全球共同遵循TCP/IP协议,才有Internet的今天。数字图书馆技术还没有这样一整套公认的符合标准规范的、赖以运作的标准协议,因此信息技术标准的选择和参与制订,对每一个数字图书馆建设者来说都是至关重要的。

五、信息传递网络化

有专家形容数字图书馆是一个“互联空间”或“电脑化空间”。数字图书馆已远远超出了传统图书馆界定的场所，它是通过互联计算机网络，把分散在世界各地的网上资源有效地连接起来，超越了时空的约束，使用户能在网络所及的任何时候、任何地点和用多种方式获得所需文献信息资源。高速、可靠的数字通信网络是数字图书馆的存在基础，数字图书馆依附于网络而存在，对内的业务组织和对外的服务都是以网络为载体，得益于网络也受制于网络。通过网络形成与平台无关的各种文献信息数字化资源集合，实现异构数据库之间、服务器之间、工作站之间的互操作性和分布式数据的网络化存取。只有最大限度地利用网络，才能最大限度地发挥数字图书馆的作用，实现数字化信息资源的全球共享。目前，在Internet上查阅信息，需要逐个站点地查询，实现数字图书馆以后，读者只要提供某个检索点，计算机就会按统一的用户界面提供所需的全部信息。

六、广泛的可存取性

数字图书馆是在网络环境下运行的一个超大型信息系统。也就是说，数字图书馆可与地区网、国家网、国际网互联，使用户可获得馆内外的大量信息资源。广泛可存取性包含两层意思：一是资源来自世界各地的政府、大学、研究机构、企业、团体和个人；二是不论什么人都可在网络所及的任何时候、地点获得所需资源。外国人称为“信息存取自由化”。因此，网络不仅是数字图书馆具有广泛存取性的基础，也是数字图书馆的核心特征之一。这是传统图书馆与数字图书馆最大的差别之一。

七、资源共享性

资源共享是传统图书馆苦苦追求的目标，但由于观念、条件和环境等许多因素制约，在传统的图书馆条件下，不可能实现真正的资源共享。而在今天的网络环境下，人们可借助网络，各图书馆都可相互交换数字化的馆藏，包括机读目录、电子出版物，用户也可以自由交换各自的数字化资源。所以数字图书馆的信息资源可包括想要包括的世界信息资源，成为全世界共享信息的资源库，这是理想化的数字图书馆特征。从目前的情况来看，由于国家利益、集团利益、版权等等一系列问题，仍然阻碍着实现真正“信息资源”共享的美梦。

八、开放性

从图书馆的角度来说，要实现信息资源共享，除技术条件外，更主要的是开放，即图书馆对所有人开放，除了要保护的一些特殊资源外，馆藏应完全对外开放。所以有人预言，未来的图书馆是一种“无围墙的图书馆”，即通过网络条件，对所有社会大众开放。有了这个意识和行动，数字图书馆的资源共享特征才能美梦成真。

九、信息提供的知识化

与传统图书馆不同，数字图书馆已经并将实现由文献的提供向知识的提供的转变。数字图书馆将图书、期刊、照片、声像资料、数据库、网页、多媒体资料等各类信息载体与信息来源，在知识单元的基础上有机组织并链接起来，以动态分布的方式为用户提供服务，而自动标引、元数据、内容检索、不同数据库的互联等知识发现与组织的技术，将成为数字图书馆的技术关键。数字图书馆信息提供的知识化，将为广大读者建立起“学术银行”、“数据仓库”，同时图书馆员也将成为知识导航员。通过信息分析和重组，形成符合用户需求的知识或帮助用户找到解决方案，由信息提供的多次满足转变为信息提供的一次满足。