

6

6

北京圖書館 出版社

数字图书馆研究

李冠强 著

北京图书馆出版社

图书在版编目(CIP)数据

数字图书馆研究/李冠强著. - 北京:北京图书馆出版社,2002.8

ISBN 7-5013-1922-7

I. 数… II. 李… III. 数字图书馆 - 研究 IV. G250.76

中国版本图书馆 CIP 数据核字(2002)第 052747 号

书名 数字图书馆研究

著者 李冠强 著

出版 北京图书馆出版社(100034 北京市西城区文津街 7 号)

发行 (010)66126153 **传真** (010)66174391

E-mail Btsfxb@publicf.nlc.gov.cn

Website www.nlcpress.com

经销 新华书店

印刷 北京双桥印刷厂

开本 850×1168 毫米 1/32

印张 5.594

版次 2002 年 8 月第 1 版 2002 年 8 月第 1 次印刷

字数 130 千字

印数 1 - 3000

书号 ISBN 7 - 5013 - 1922 - 7/G·516

定价 16.00 元

目 录

第一章 数字图书馆及其相关概念	(1)
第一节 数字图书馆产生背景	(1)
第二节 数字图书馆的定义、特点及其相关概念	(10)
第三节 数字图书馆的组织模式	(23)
第四节 数字图书馆与电子图书馆等相关概念的比较	(29)
第二章 数字图书馆网络系统建设	(35)
第一节 数字图书馆网络系统建设研究	(35)
第二节 数字图书馆的关键技术研究	(42)
第三节 数字图书馆智能化建设的项目需求	(47)
第四节 数字信息归档系统	(51)
第五节 多媒体高速数据网络体系	(57)
第六节 数字图书馆用户界面研究	(59)
第三章 数字图书馆的信息资源建设	(71)
第一节 数字图书馆的信息资源建设	(71)
第二节 数字信息制作、检索与保存	(90)
第四章 数字图书馆的网络信息安全	(107)
第一节 网络信息安全现状与防范机制	(107)
第二节 数字图书馆的网络信息安全	(111)
第五章 数字图书馆与传统图书馆的关系	(118)
第一节 数字图书馆与传统图书馆的区别与比较	(118)
第二节 数字图书馆对信息服务的影响	(124)

第三节	数字图书馆是否能完全取代传统图书馆·····	(128)
第六章	数字图书馆对社会发展的影响·····	(133)
第一节	数字图书馆的社会价值·····	(133)
第二节	数字图书馆的效益分析·····	(134)
第三节	数字图书馆与文化教育·····	(142)
第七章	中外数字图书馆的发展现状·····	(147)
第一节	数字图书馆的研究历程·····	(147)
第二节	各国数字图书馆项目概况·····	(150)
第三节	中国数字图书馆的发展现状·····	(170)

第一章 数字图书馆及其相关概念

谢拉曾经说过：“图书馆如果不是立刻反映，也是最终反映着社会的变革。”纵观图书馆的发展史，从古代、近代到现代图书馆，文献载体经历了由甲骨到帛、纸，又到缩微胶片的几次重大变革，图书馆由古代的藏书楼发展到今天为全社会所利用的各类公共图书馆、高校图书馆、研究图书馆等。图书馆的每一次进步无不反映着社会生产力水平的提高和文明的进步。计算机的出现与普及，通讯手段的发展，以及两者结合而产生的网络技术，无不在昭示着图书馆将面临着又一次飞跃，即由现代图书馆发展成为在网络环境下生存的数字图书馆。

第一节 数字图书馆产生背景

20 世纪末，随着计算机技术、数字存储技术、现代通讯技术、网络技术的飞速发展与有机结合，数字信息日益广泛地用于综合信息资源。像其它格式的信息一样，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。每当人们谈到图书馆时，首先想到的是雄伟的建筑、汗牛充栋的书库和宽敞的阅览室，但在数字信息浪潮的冲击下，一种新型图书馆——数字图书馆正应运而生，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信息高速公路的

多媒体信息管理系统下,数字图书馆的信息不再存放在一幢物理建筑里,而是分布在世界各地的网络系统中,呈现出一种虚拟形态。图书馆也从单纯获取知识的场所变成一个集获取知识、休闲娱乐和相互交流于一体的多功能场所。数字图书馆的出现将对人们的生活方式和社会发展产生巨大影响。

1. 数字图书馆产生的历史背景

回顾 20 世纪的 100 年中,冲击和影响人类社会生活各个方面幅度最大、速度最快、震撼最为强烈的,莫过于延续至今仍在飞速发展着的信息技术革命,它使每一个人都深刻地感受到自己正经历着全方位的急剧变革,其影响涉及到社会、政治、经济、科技、文化、思想和价值观念等几乎所有方面。

在这种形势下,图书馆也同样毫不例外地经受着信息技术革命浪潮的冲击和影响。问题还不仅仅如此,从历史上看,信息技术水平在很大程度上被视为划分图书馆发展阶段的重要依据。比如,以泥板、莎草纸、甲骨、竹简等为信息载体,以手抄为记录手段的时代,只能造就以收藏保管为主的古代藏书楼式的图书馆;而造纸和印刷术的发明与发展促使古代藏书楼走向解体并孕育了近代图书馆,从而为现代图书馆的诞生奠定了基础;20 世纪 60 年代是现代图书馆诞生的重要发展阶段,计算机在图书馆的应用为图书馆进入自动化阶段揭开了序幕。从总体上看,20 世纪 60 年代以前的图书馆是以收藏印刷品为主的纸张图书馆,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确实已经存在的文献,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、通讯以及网络为核心的现代信息技术开始全面渗透进图书馆领域,将图书馆推向自动化发展

阶段。特别是 1969 年美国国会图书馆开始正式发行 MARC II 机读目录磁带,为实现图书馆大部分业务工作的自动化奠定了基础,开创了使用机读目录的新时期,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。

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并与各馆建立的自动化集成系统连成网络。在这个阶段,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,还建立了机读目录数据库和二次文献数据库,用户可以通过联机公共存取目录(OPAC)查询馆藏书目,通过联机检索终端查找二次文献信息。

从图书馆自动化发展历程来看,一直强调图书馆内部业务工作的自动化和检索工具编制与使用的自动化,而不是图书馆馆藏文献本身存储与检索的计算机化。但是,到了 20 世纪 80 年代末和 90 年代初,得益于计算机技术、通信技术、网络技术、高密度存储技术、多媒体技术等新技术的高速发展、不断突破及其有机结合,图书馆自动化开始进入高级发展阶段,向着高度自动化、电子化、网络化、虚拟化的深度和广度进军。

近 10 年以来,随着信息技术的长足进步和飞速发展,以及国际互联网的建立和广泛普及应用,在全球建设信息高速公路热潮的直接带动下,国际图书馆界对许多新名词、新术语、新概念、新系统表现出极大的热情和兴趣,如“数字图书馆”、“电子图书馆”、“虚拟图书馆”和“无墙图书馆”等,这些内容已成为众多报刊杂志、广播电视以及各种学术会议的热门话题。

1993 年,在德国的埃森召开了首届国际电子图书馆会议。

1994年,在美国德克萨斯召开了首届国际数字图书馆会议。美国计算机协会和美国信息科学学会及其它一些著名学会、协会的会刊都出版了与数字图书馆有关的专辑。1996年3月,美国计算机协会信息检索专业组(ACM SIGIR)、美国电气与电子工程师学会(IEEE)、美国信息科学学会(ASIS)等几大学术组织在 Bethesda 召开了规模很大的首届 ACM 数字图书馆国际会议。

2. 数字图书馆产生的技术背景

20世纪80年代以来,光电技术开始得到广泛应用,光纤通信已经逐渐成为通信传输的主要手段,千百万甚至上亿人在同一个通信网同时通信,万亿次速度巨型计算机、千百万台计算机联网运行的时代正在到来。

经过20多年的发展,当今世界上最大的国际网 Internet 近年来几乎以席卷全球的态势在全世界各个角落被广泛地应用。如今的 Internet 已经成为建立于一种高度灵活的通讯技术之上的全球数字化信息库,提供了用以创建浏览、访问、搜索、阅读、交流信息的形形色色的服务。Internet 将信息库中的信息以数字化形式记录、编辑、组织成菜单,作为超媒体文档文件存储或者保存在文本之中。此外,能够通过数字化信息库进行访问的信息,可以由那些被即时地搜集、交流并提交而无需存储的数据(包括声音和视像)所构成。再者,由于各种服务已被集成化并建立了交叉参照,因而用户可以“无缝地”从一台计算机上的信息链接到另一台计算机上的信息,并且可以将一种服务上的信息链接到另一个服务上的信息。随着 Internet 功能的不断强大,促使它迅速地发展。1993年,Internet 网上的主机数量为180万台,进入21世纪时,Internet 入网数已达到1.2亿台,用户已突破10亿大关,基于 Internet 的

产品与市场已达到 200 亿美元。

随着现代信息技术的发展,各种电子出版物大量出现。电子期刊被视为一种经过正常编审,以电子形式出版并有足够存储设备的出版物。实际上电子期刊是一个时刻在扩充、变化的活的数据库。电子期刊出版速度比普通印刷型期刊快得多,它利用计算机进行阅读、编辑、修改、定稿、拷贝复制。联机的电子出版物可以通过电子邮件直接向编辑部发送原文,读者可以直接在终端显示器上阅读、打印,操作极其方便。

多媒体技术是把文字、声音、图像三种不同的媒体结合起来,使计算机能同时处理和综合控制三种媒体集成信息系统的现代化技术。多媒体将实现人们的远距离购物,召开电子会议等设想。多媒体最大应用领域是出版市场,其中电子出版物的出现是多媒体进入图书馆的一个重要方面,它使图书馆馆藏结构发生了重大变化。

数据库技术是对信息资源的整理、保存、管理和使用的最有效的手段,是提高管理水平和效率的核心。随着现代科学技术和信息产业的发展,世界各国都竞相发展不同用途、类型各异的数据库,其内容被广泛扩展到科技、经济、管理、社会生活、教育等各个领域。在日益信息化的时代,数据库愈显示出它的重要性。可以说,当今社会,一个数据库开发利用的程度,是这个国家信息化程度的重要标志。各个系统的数据库通过卫星、光纤网络与世界各地联系在一起,在全球信息交流中发挥着重要的作用。Internet 是最大的信息库,其网上的数据库多达上万个,包括从美国国会图书馆到欧洲各国的公共图书馆和大学图书馆的几百个联机目录和大量数据库。

3. 数字图书馆产生的理论背景

早在 20 世纪 80 年代,托夫勒在其闻名于世的代表作《第三次浪潮》中指出:第一次浪潮是指几千年来缓慢发生的农业革命;第二次浪潮是指产业革命,在与第一次浪潮时代的旧体制不断冲突中,第二次浪潮逐渐渗透到社会的各个方面;第三次浪潮当时并未给出特定的称谓,是指未来时代的信息革命。纵观近代图书馆的发展历程,我们可以仿照托夫勒的说法,把对图书馆界产生巨大影响的三次技术革命叫做图书馆发展的三次浪潮:

——第一次浪潮:中世纪和近代创立的各种分类法,特别是 1876 年美国杜威编制的《十进分类法》。

——第二次浪潮:始于 20 世纪 60 年代的计算机革命,如机读目录 MARC 的出现,联机信息检索系统等。

——第三次浪潮:指目前正在进行的数字图书馆革命,数字图书馆承担着把人们从第二次浪潮的束缚中解脱出来的使命。

与图书馆发展的三次浪潮相对应,图书馆的发展史上也出现了三个发展阶段。国内外许多图书馆和文献信息学科的专家一致认为这三个发展阶段是:纸媒体图书馆、自动化图书馆、数字图书馆。而发达国家的图书馆自动化历程也有三个发展阶段,第一阶段可称为图书馆自动化发展的初级阶段,即图书馆自动化管理集成系统发展阶段,以 20 世纪 60 年代美国国会图书馆正式发行 MARCII 型的机读目录为标志,它在北美得到广泛应用,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阶段。第二阶段为图书馆在网上进行全球性、整体化的电子文献信息服务的新阶段。这一阶段发生在 1985 年左右,以 CD-ROM 光盘和局域网络开始在图书馆得到应用为主要标志,使人

们开始可以在图书馆、办公室、实验室甚至家中访问图书馆的书目机读目录、单位局域网上的光盘数据库和大型文摘社的检索系统,使 70 年代出现的大型文献信息中心充分发挥了效益,特别是 90 年代信息高速公路的迅猛发展,将图书馆网上的电子文献信息服务推向了全球性服务的新阶段。第三阶段是图书馆自动化的高级发展阶段,也称为数字化图书馆阶段。

这三个阶段并不是截然分割的,它们是相互渗透的,前一个阶段又为后一个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,它实质上是图书馆馆藏目录的数字化,它们构成机读公共目录 OPAC,是第二阶段电子文献信息服务的重要组成部分,也是在进入数字化图书馆阶段中,有一部分文献信息载体还没有被数字化的资料的重要补充,可以通过它们得到这些资料的索引、摘要等。在第二阶段中,已对数字化图书馆的技术做了大量研究,出现了一些数字图书馆的雏形,如在 Internet 网上的许多 WWW 服务器,已可以提供文本、图形、声音、运动图像和多媒体资料查询。1988 年美国自然科学基金会提出研究电子图书馆的“水星计划”(Mercury Project),以及 1994 年扩大为六个学校的新一轮数字化图书馆创始计划。1990 年开始,美国国会图书馆进行的“美利坚记忆”(American Memory)的电子图书馆创始计划,这些都是为研制数字化图书馆在理论上、技术上和实际应用各种政策及措施上做了准备。第二阶段中,对各种文献载体的数字化,大容量数据的存储和管理技术,各种文献载体的访问及传输技术等又做了大量的研究,为数字图书馆技术走向实用化做了准备。

20 世纪 60 年代电子书刊的兴起,尤其是以后电子检索期刊在国际联机检索上的成功应用,导致了电子图书馆的迅速发展。

美国在 1962 年就在西雅图举办过“21 世纪图书馆”的展览,当时就提出了“没有图书的图书馆”的观点,这是数字图书馆最先的舆论准备。1975 年,美国图书馆学家克里斯蒂安(R. W. Christian)出版了《电子图书馆:书目数据库:1975——1976》一书,书中首次提出了“电子图书馆”这个名词。随后,美国著名图书情报学家兰开斯特(F. W. Lancaster)在其《走向无纸情报系统》和《电子时代的图书馆与图书馆员》两部著作中都描绘了电子图书馆及数字图书馆的前景。

1983 年美国科莱因(H. F. Cline)和辛诺特(L. T. Sinnott)在专著《电子图书馆——自动化对学术图书馆的影响》一书中频繁使用了“电子图书馆”这个术语。

在日本,1985 年日本邮政省首先提出电子图书馆的构思,东芝公司甚至开发出了“电子图书情报系统”。同年 11 月,日本 40 多家主要新闻媒介和出版社发起成立了“电子图书馆研究会”,不久,日本的商业电子图书馆也投入了使用。

我国对数字图书馆的研究也有近 20 年的历史,前期主要是介绍国外的译文和译著。1985 年,由郑登理、陈珍成译校的兰开斯特《电子时代的图书馆与图书馆员》,由科技文献出版社出版。1988 年,刘建明节译了由美国图书馆协会所属图书馆与情报技术协会主席道林(Kenneth E. Dowlin)的著作,以《电子图书馆》为题,登载于上海的《图书馆杂志》。同年,还有乔欢等翻译的英国情报学家詹姆斯·汤普森(Jeames Tompson)所著的《图书馆的未来》,由书目文献出版社出版。1991 年赖茂生发表了《电子图书馆的构想与实现》一文,较早地介绍了电子图书馆运行的构想。早期卓有成效的研究者当推上海交通大学的杨宗英教授,1993 年,他在《大学图书馆学报》以《电子图书馆的崛起》为题,连续发表了数篇有关电

子图书馆的力作,为后期的学术研究和建设奠定了基础。他认为,现代电子技术的飞速发展,计算机技术和通信技术的普遍应用,国际书刊市场上电子书刊的发行,是电子图书馆发展的社会基础和物质基础。近年来,中国科学院文献情报中心徐引篪、霍国庆所著的《现代图书馆学理论》,武汉大学图书情报学院黄宗忠教授的《论 21 世纪的图书馆》、《论 21 世纪的虚拟图书馆与传统图书馆》,中科院文献情报中心的汪冰博士的《电子图书馆理论与实践研究》、《电子图书馆及其相关概念辨析》,北大刘兹恒先生的《试论虚拟图书馆与传统图书馆的关系》等一批论述电子图书馆、数字图书馆的力作相继问世,提出了令人信服的理论观点和未来发展模式的思路。

4. 数字图书馆产生的实践背景

数字图书馆是图书馆自动化、网络化的产物,它向上可追溯到早期的图书馆自动化系统,世界上最早的计算机图书馆检索系统是 1954 年由美国海军兵器中心图书馆建立的,它是世界上第一个文献检索系统。随后图书馆自动化在美国兴起,20 世纪 70 年代美国建立了三大图书馆网络:联机计算机图书馆中心(OCLC)、研究图书馆信息网(RLIN)、西部图书馆网(WLN)。目前来说,OCLC 已成为世界最大的图书馆网,成员馆达 1 万余所,一年收录书目近 200 万条,现在书目已达 3000 万条之多。

近年来,随着计算机技术、网络技术的飞速发展,尤其是 Internet 的迅猛发展,进一步推动了图书馆自动化系统向数字化、网络化发展。在这种形势下,国外发达国家加紧对数字图书馆的研究与开发,当前国外比较重大的数字图书馆研究计划有:英国国家图书馆“争取创新”计划,美国国家科学基金会和国防高级研究

计划局所进行的数字图书馆计划等。

我国在研究建造数字图书馆方面也正在做自己的努力。首先,我国的社会信息化建设步伐不断加快,有着重大进展。1993年9月我国建成新的分组交换数据网(CHINAPAC)。1994年10月,中国数字数据网(CHINADDN)开通,我国开始与因特网连接,目前已有中国科学院中国国家技术与设施网(NCFC)、中国科学院高能物理研究所网(IHEPNET)、中国教育科研网(CERNET)、中国公用计算机互联网(CHINANET)、金桥网(GBN)等子网与因特网直接联网。

我国图书馆网络也在积极建设中,有的已取得了显著进展。1991年以中山图书馆为中心的珠江三角洲公共图书馆自动化网络计划开始实施,至1994年5月已建成三级结构的地区图书馆网,上通国家图书馆网,下连市县图书馆,基本覆盖了珠江三角洲地区。连接教科网的中科院、北大、清华文献信息共享网络(APTLIN)建设工程业已起动,其近期目标是实现公共书目检索、联机编目和馆际互借管理。

中国国家图书馆从1995年开始跟踪国际数字图书馆发展,1996年初申请成立数字图书馆实验项目,并由国家计委批准该项目为“中国实验型数字式图书馆”,1998年又由国图带头,倡议实施“中国数字图书馆”工程,极大地推动了我国数字图书馆的建设与发展。

第二节 数字图书馆的定义、特点及其相关概念

1. 数字图书馆的定义

要谈数字图书馆,首先要了解什么是数字化,这就要从二进制

的数字信号说起。说起数字信号并不神秘,早在发明电报之时,它就被派上了用场。通过电报滴答声的两种物理状态代表数字的 0 和 1,再用 0 和 1 的组合编码来代表字母 A—Z 和十进制数 0—9,以及其它字符。在信息传递过程中,发报方把发送内容逐字译成 0、1 数字码,通过脉冲电信号传递出去;接收方通过收到滴答声表达的数字信息,再译出电报内容。由此可以看出,0 和 1 的二进制数字信号,可表达和传递内容复杂的信息。

在信息传递和处理领域,二进制数字信号是现实世界上最容易被表达,物理状态最稳定的信号,而且抗干扰强,传输的信息内容不衰减。正是由于二进制信号的这些优越性,各种电子信息传递媒体,如广播、电影、电视、通讯等领域,纷纷把信息加工与传递技术采用数字方式处理,形成了一股信息数字化的热潮,这就是我们所要说的“数字化”

计算机的出现是信息数字化时代的开始。冯·诺伊曼式计算机的重要特征之一就是使用二进制,计算机内部处理信息时只识别 0、1 数字信号。所有要处理的信息,都是通过输入设备或是模/数转换设备把它们自动转换成二进制的电信号,然后以编码方式在机内对信息处理完毕后,再通过输出设备或数/模转换设备把它转换成人们能够识别的形式输出。数字化实质上是计算机技术在信息领域的全面应用。

多媒体技术的开发,更使数字化应用全面铺开。所谓媒体,在计算机领域有两种含义,一是指用以存储信息的实体,如磁盘、光盘和半导体存储器等;一是指信息的载体,如数字、文字、声音、图形和图像等。多媒体技术中的媒体是指后者,多媒体实际上是一种二进制的数字媒体,它是通过模/数转换技术,把各种模拟信息,如文本、声音、图像等各种媒体转换成二进制数字信号进行存储、

传输或加工处理的。这种方式彻底克服了其它媒体在使用模拟信号时信号抗干扰性差、易衰减等弊端。

信息技术的进步,促动图书馆自动化从业务管理全面走向文献信息资源的综合管理。在此转化当中,计算机性能的提高,对图书馆的数字化起了决定性的作用。图书馆的文献资源全部以计算机可读形式,即数字化形式存放,或者是存放在光盘里,或者是存放在磁盘里,或者是存放在硬盘里,对外服务时或采用联机(在线)调用形式,或采用制品型的外借方式,这样,便推动了数字图书馆的产生与发展。

关于数字图书馆的定义有许多说法,广义而言,一个数字图书馆是计算机可处理信息的集合或此类信息的一个存储处,许多学者据此对数字图书馆的定义做出了各自的解释,这些定义大致有如下4种不同的出发点:

(1) 有些研究者将数字图书馆等同于计算机化的、网络化的图书馆系统。他们用这一术语来描述可通过因特网存取数据库或 CD—ROM 信息产品,特别是包含书刊文献完整内容的产品。

(2) 有些研究者认为,所谓的数字图书馆只不过是一种新的信息技术在图书馆和类似机构中的应用。如英国图书馆的 G. Jefcoate 即认为,数字图书馆作为一个已广泛被人接受的术语,意指图书馆使用数字技术来采集、存储和保存信息并提供存取信息。他所在的图书馆研究与创新中心已启动了 20 个与数字技术应用相关的项目。该馆相信,他们所创建的数字图书馆将由大量重要的数字化馆藏文献组成,将以一种组织有序的服务框架,在任意时间向任意地方的任意用户提供需求服务。

(3) 有些研究者把数字图书馆视为一种新类型的图书馆。如美国情报科学学会主席 M. Buckland 教授在《重新设计图书馆服