

XIANDAI XINXI WANGLUO SHIYONG JISHU JIANGZUO

现代信息网络实用技术讲座 第二辑

数字 图书馆

● 王忠华 周 勇 编 著



北京邮电大学出版社
www.buptpress.com

现代信息网络实用技术讲座

数字图书馆

主 编：李 勇 赵呈领 陶智慧
副主编：陶智勇 乌文全 贺 钢
陈 健 姚永清 汪顶武
编 著：王忠华 周 勇

北京邮电大学出版社

·北京·

内 容 简 介

数字图书馆是一个新兴的、涉及到互联网、多媒体、数据库、信息安全等方面技术的计算机应用领域。本书介绍了数字图书馆的概念、数字图书馆的基本原理与技术、文献分类与数据描述标准、数字图书馆的建设探索,以及数字图书馆的研究与应用实例等内容。

本书的读者对象是广大的数字图书馆的爱好者、使用者,也可供计算机、图书馆等领域的研究人员参考。

图书在版编目(CIP)数据

数字图书馆/王忠华,周勇编著. —北京:北京邮电大学出版社,2002
(现代信息网络实用技术讲座)

ISBN 7-5635-0595-4

I. 数... II. ①王... ②周... III. 数字图书馆—基本知识 IV. G250.76

中国版本图书馆 CIP 数据核字 (2002) 第 021974 号

书 名: 数字图书馆

作 者: 王忠华 周 勇

责任编辑: 蒋 亮

出 版 者: 北京邮电大学出版社(北京市海淀区西土城路 10 号) 邮编: 100876

发行部电话: (010) 62282185 62283578 (传真)

经 销: 各地新华书店

印 刷: 北京源海印刷厂

开 本: 850 mm × 1168 mm 1/32 印 张: 3.875

印 数: 0 001—3 000 册 字 数: 101 千字

版 次: 2002 年 5 月第 1 版 2002 年 5 月第 1 次印刷

ISBN 7-5635-0595-4/TN · 269 全套定价: 40.00 元 本册定价: 10.00 元

如有印装质量问题请与北京邮电大学出版社发行部联系

《现代信息网络实用技术讲座》

总 序

在知识经济初见端倪和网络的浪潮涨潮落的世纪之初，世界的经济体系更加依赖于人类的知识和智慧。人们更加需要信息，并且需要随时随地地获取信息。以因特网技术为核心的现代信息网络从根本上影响着人们获取信息的方式，因而，也进一步的影响了人们的工作、学习、消费、休闲、交往等生活方式。

人们迫切需要高速实时的现代信息网络。现代信息网络就是城市、国家或国际的信息基础设施（俗称“信息高速公路”）。因为只有具备了这个平台才有可能加速传播和共享知识，大大降低知识传递的成本，开创新的工作方式、管理方式、商贸方式、金融方式、人际交往方式、文化教育方式、医疗保健方式、消费与生活方式以及休闲方式。使知识经济真正成为数字经济和网络经济。

近年来人们一直在探讨现代信息网络的未来模式与发展趋势。这一切仁者见仁，智者见智，尚有很多的争议。但有一点可以肯定：现代信息网络要对全社会一般家庭成员的生活方式产生一般性的影响或者使之发生根本性转变，仅仅靠局部的个别技术和少数人的参与是不够的。网络购物在中国为何举步难艰，原因很清楚，我们缺少配套的其他信息技术以及百姓的认同意识。比如中国的信用卡持有率很低、通用性很差，邮政服务不尽人意，商用数据库信息匮乏等等。

现代信息网络的技术必须向百姓普及，必须为百姓所接受，其技术及应用已经不再只是信息技术行业从业人员关注的话题，它已深入到当今社会越来越广泛的人群之中，并被他们所应用，与我们的生活、工作越来越息息相关。现代信息网络的技术及其

应用作为实现社会信息化的一项重要支持技术和操作平台，其发展、应用和普及尤其令人瞩目，也受到世界各国政府的广泛重视。人们对现代信息网络的其技术及应用了解的需求也日益突出。

目前，有关现代信息网络方面的图书数量尽管很大，但大多内容庞杂、面面俱到、理论深奥、价格不菲，难于为广大普通百姓所接受，更谈不上向这类人群普及。为了帮助大家及时了解信息技术的发展、掌握新技术的应用方法，为顺应网络通信技术的发展大潮和网络通信关注者的迫切需求，北京邮电大学出版社周明老师与我们共同策划编写了这套《现代信息网络实用技术讲座》丛书，供大家学习使用，希望为广大关注现代信息网络的技术领域的普通百姓提供专业的技术咨询和实用的信息服务。这套丛书紧密结合实际。重点介绍近年来迅速出现并发展起来的新技术、新应用。主要内容有：IP 电话、电子商务、视频会议系统、智能大厦、数码相机、网上教育、视频点播、数字图书馆等。丛书的特点是结合发展，全面介绍新技术、新概念、新应用，突出实用性、系统性，内容详实、丰富，便于读者理解和掌握。相信不管是普通百姓还是专业人士都会开卷有益。

这一套丛书的宗旨在于：引导大众迎接信息网络时代的生活，以通俗的方式普及信息技术。纵览这套丛书的作者，他们其中有华中师大的博士和中青年教师、华中科技大学的博士和中青年教师、武汉邮科院和省市电信公司的高级工程师，他们把自己对现代信息网络的深刻理解和实际经验，凝聚在这套现代信息网络实用技术讲座中，奉献给大家。这些中青年才俊，他们为了整个文稿的简捷、通俗易懂而不厌其烦，几易其稿，这令我们既感动又宽慰，北京邮电大学为这套丛书的出版倾注了大量的精力，我们谨以此致以诚挚的谢意。殷切希望广大读者和各有关方面提出宝贵意见和建议，以便这套丛书日臻完善。

《现代信息网络实用技术讲座》丛书编委会

前 言

随着现代信息技术的深入发展,信息载体的数字化以及信息传播的网络化发展趋势,使以书刊资料为主要收藏载体的传统图书馆面临巨大的挑战,逐渐难以适应数字时代的要求,因而图书馆的数字化是一个必然趋势。数字图书馆的出现给图书馆事业的发展带来新的契机,数字图书馆正成为全球范围内信息基础设施建设的热点领域,也将成为 21 世纪图书馆事业发展的主旋律。

本书全面介绍了数字图书馆的基本概念、原理、技术与应用,其主要内容包括数字图书馆的基本概念、数字图书馆的基本原理与技术、数字图书馆的建设探索、数字图书馆的研究与应用实例等。其中第一章介绍了数字图书馆的基本概念及国内外的发展状况;第二章论述了数字图书馆的基本原理及其关键技术;第三章介绍了数字图书馆所涉及到的相关标准;第四章从数字图书馆建设的角度讨论了数字图书馆的解决方案、资源建设、面临的技术挑战等;第五章介绍了国内外数字图书馆的研究状况及其应用实例,以便读者对数字图书馆的使用有一个基本的了解。

数字图书馆是一个新兴的领域,而且其覆盖面非常广泛,加上编者水平有限,书中难免有不当及误漏的地方,敬请读者批评指正。

目 录

第一章 数字图书馆简介	1
1.1 数字图书馆的基本概念	1
1.2 数字图书馆的功能与作用	10
第二章 数字图书馆的基本原理与技术	13
2.1 因特网与万维网	13
2.2 数字图书馆的基本结构与组成	22
2.3 数字图书馆的基本原理与技术	24
2.4 信息检索技术	30
2.5 数字图书馆涉及的其它问题和技术	37
第三章 文献分类与数据描述标准	41
3.1 文献及其分类标准	41
3.2 描述性元数据及其标准	46
3.3 XML 及其标准	54
第四章 数字图书馆的建设探索	59
4.1 数字图书馆的解决方案概述	59
4.2 几种数字图书馆的解决方案介绍	61
4.3 数字图书馆的资源建设	66
4.4 我国数字图书馆的建设情况	70
4.5 数字图书馆建设面临的技术挑战	78
第五章 数字图书馆的研究与应用实例	83
5.1 数字图书馆研究	83
5.2 数字图书馆应用实例	97

附录一 国外数字图书馆相关资源站点	111
附录二 国内数字图书馆相关资源站点	113

第一章 数字图书馆简介

1.1 数字图书馆的基本概念

随着 Internet 在世界范围的迅速发展, 信息资源的全球交流与共享正在变成现实。信息存储和检索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。Internet 对传统图书馆工作提出了严峻的挑战, 同时也带来了新的发展机遇。

1.1.1 数字图书馆的起源与发展

1. 数字图书馆的起源

1945 年 7 月, 当时担任美国科学研究和开发办公室主任的 Vannevar Bush 在《大西洋月刊》(Atlantic Monthly) 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存储、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信息机“Memex”, 这一直被认为是数字图书馆概念的起源。Bush 设想的 Memex 系统以缩微作为信息的存储体, 计算机负责信息的索引和查询, 实际上是一个计算机辅助检索系统。二十年后, 试图用数字计算机改变图书馆面貌的美国学者 Licklider 于 1965 年出版了《未来的图书馆》

(Library of future)一书，大胆地预测了20世纪末数字图书馆的面貌。以上两位学者被认为是数字图书馆的先驱。

第一次真正应用计算机存储和处理图书馆信息的尝试可追溯到20世纪60年代末，首先是从建立法律和科学文献的全文索引开始的，美国俄亥俄法律自动研究系统能够提供法律状态的联机检索，空军建立的法律信息电子系统(LITE)索引了当时美国所有的法律法规和司法解释，提供检索服务。70年代IBM基于文本存储与检索技术开发了STAIRS系统，安装在许多大型机中，提供给当时许多图书馆用于流通管理等。

同时早期产生了一些信息服务，如联合编目、法律信息系统和美国国家医学图书馆的MEDLINE联机数据库系统等，这些系统要求用户训练有素，通过低速的通信联接与中央计算机交换信息。这种模式一直沿用到80年代中期，这时大型国际联机检索系统，例如DIALOG、ORBIT、STN等已获得普遍应用，通过它们，MEDLINE、化学文摘(CA)、科学文摘(INSPEC)等非常重要的科学数据库得到了广泛传播。

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(NII)和因特网的迅速普及处在同一时期。数字图书馆的基础根植于整个80年代对联机情报检索的追求和探索，以及全文本、多媒体信息处理技术的成熟，其发端可归因于因特网出现后美国政府对信息基础设施的研究和投入。

2. 数字图书馆的发展

20世纪60年代，计算机在图书馆的应用为图书馆进入自动化阶段揭开序幕。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透到图书馆领域，将图书馆推向自动化阶段。而在70~80年代，计算机与通信相结合，又促进了图书馆联机系统和网络化的发展，一批联机编目网络和一些商业性联机检索系统相继出现，并进一步发展在各馆之间建立起互连网络。到了80

年代末和 90 年代初,得益于计算机技术、通信技术和网络技术、高密度存储技术、多媒体技术等高新技术的高速发展,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。进入 90 年代以来,随着 Internet 的建立和广泛普及和应用,在全球建设信息高速公路热潮的直接带动下,数字图书馆这一概念逐步得到社会的认可,并成为当今的热门话题。

数字图书馆是伴随国际互联网的发展和普及而兴起的,大致发端于九十年代初。美国于 1991 年率先开始研究数字图书馆。1992 年 7 月,美联邦政府提出了《信息基础设施建设与科技法案》,在这一法案中明确界定了数字图书馆的含义、功能和技术应用的大体范围。将数字图书馆研究及建设纳入其国家信息基础建设(NII)的重要组成部分,以六所大学作基地,以试验数字化信息的收集、组织、存储并传输。其目的就是希望通过数字图书馆这种国家信息基础设施将知识的高速公路铺向每个美国人的家庭。

美国数字图书馆的研究和建设走在世界各国的前列。并且,对于建设数字图书馆的意义也日益重视。美国政府“国家计算、信息、通信指导办公室”(NCO for CIC),自 1994 年以来每年发表一本“蓝皮书”,这是一种正式的重要研究与发展的官方报告,近年来,报告对数字图书馆战略意义的描述越来越重视。

1995 年的蓝皮书“用于国家信息基础设施的技术”列出了九项 NII 应用的国家级挑战,依次为:数字图书馆、危机及应急管理、教育及终身学习、电子商务、能源管理、环境控制和废品最小化、保健、加工处理和产品、公众存取政府信息。报告指出,数字图书馆是无墙的知识中心的基础,并强调数字图书馆技术将被用于所有其它的国家级挑战应用之中。

1996 年蓝皮书“高性能计算和通信:用于未来美国信息的基础”介绍了 18 个在美国进行的数字图书馆项目状况及 IP

地址。

1997 年蓝皮书“推进信息技术的前沿”明确将数字图书馆列入有效技术之中。

1998 年蓝皮书“用于 21 世纪的技术”则又将数字图书馆在 CIC（计算、信息、通信）的六个研究发展重点项目中名列于第一位。

在美国之后，其它许多国家也投资研究建造自己的数字图书馆，如加拿大、英国、法国、意大利、荷兰等。例如，加拿大政府在 1996 年 5 月公布的《建设信息社会：使加拿大进入 21 世纪》的行动计划中，十分重视在信息内容方面的建设；1997 年遗产和工业部联合成立了“数字化工作小组”，统筹全国的数字化工作，同时还计划修改版权法，解决多媒体、因特网发展带来的知识产权问题。

我国自 1995 年开始有数字图书馆的研究文章发表，清华大学、北京大学、上海交通大学、复旦大学、华南理工大学等院校开始了数字图书馆的研究。

1996 年 7 月，我国由国家图书馆、中山图书馆、上海图书馆、深圳图书馆、辽宁图书馆、南京图书馆、文化部文化科技开发中心共同承担实施“中国试验型数字图书馆”（CPDLP）项目。该项目的实施是我国数字图书馆建设的一个开始标志，带动了我国信息基础设施建设的资源建设向数字图书馆方向迈进。

1.1.2 数字图书馆的定义与特征

数字图书馆（Digital Library，简称 DL）是当前最热门的话题之一，相关的定义有上百种之多。然而对于它的定义，人们普遍认为目前还没有一个公认的定义，这是由于数字图书馆所涉及的领域具有明显的跨学科特征。下面就列出几种有关数字图书馆的

典型定义：

“数字图书馆就是以数字形式存储和处理信息的图书馆。”

“数字图书馆是利用数字技术将各种不同载体、不同地理位置的信息资源用数字技术存储，以实现网络查询和传播的大型信息系统。”

“数字图书馆一般而言是指利用当今先进的数字化技术，通过诸如 Internet 国际互联网等计算机网络，使人数众多且又处在不同地理位置的用户能够方便地利用。”

“数字图书馆是一个数字化系统。它将分散于不同载体，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结合，提供及时利用，实现资源共享，其核心是数字化和网络化；其实质则是形成有序的信息空间。”

“数字图书馆是全球信息高速公路上信息资源的基本组织形式，这一形式满足了分布式面向对象的信息查询需要。”这个定义中的“分布式”是指跨图书馆（跨地域）和跨物理形态的查询，“面向对象”是指不仅要查到线索（在哪个图书馆），还要直接获得要查的东西（对象）。这个定义类似于说目前的图书馆”是社会信息资源的一种主要组织形式，满足了人们借阅书刊等基本信息需要。”

虽然人们给数字图书馆下了不同的定义，它们从不同的角度去认识数字图书馆。但是数字图书馆应有以下几个基本特征：

1. 数字化资源

大量的数字化资源是数字图书馆的“物质”基础。所谓数字化是指把信息用电磁介质，按二进制编码的方法加以存储和处理，把原先用纸张存储的信息转变为用计算机存储和处理的信息。数字图书馆的目的是直接提供读者所需的最终信息，而不只是二次文献（获得文献的线索），但数字图书馆并不排斥书目数据、索引文摘等二次文献，因而它们也是数字图书馆的组成部

分。由于把各种不同载体的信息数字化后统一于 0 和 1 这两个简单的数字，书籍、期刊、录音录像带、乃至古籍善本、稀世字画甚至 X 光片，都消失了原本的物理形态，只要是数字化的数据，就能被同时获取。因此，数字化是数字图书馆的一个基本特征。

2. 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。只有充分利用好网络，才能充分发挥数字图书馆的作用。

3. 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之后，数字图书馆可以实现“联邦检索”，全球数字图书馆将像现在的 Internet 一样，把全球的数字化资源联为一体，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。

综上所述，数字图书馆就是对有高度价值的文本、图像、声音、视频和科学数据等多媒体信息进行收集，组织规范性的加工，进行高质量保存和管理，实施知识增值，并提供在广域网上高速横向跨库连接的电子存取服务。同时还包括知识产权、存取权限、数据安全管理等。

1.1.3 数字图书馆的相关概念

由于数字图书馆所涉及的领域具有明显的跨学科特征。因此，目前有许多与其相关的概念出现在各种文献当中，如电子图书馆、虚拟图书馆、无墙图书馆、虚拟现实图书馆等。这些概念在某一方面意义与数字图书馆相近，但另一方面又反映了其特有

的特点。为了有助于更加准确地理解数字图书馆这一概念，下面分别介绍这些与数字图书馆相关的概念。

1. 电子图书馆

电子图书馆 (Electronic Library) 在含义上同数字图书馆最为相近。国内外一般认为，电子图书馆是数字图书馆的早期提法，1992 年 ~ 1994 年间这两个概念都在使用，1994 年以后使用“数字图书馆”的逐渐多起来。最早对“电子图书馆”这一概念给出明确定义的是美国人 K. Dowlin，他在 1984 年出版的《电子图书馆：前景与进程》一书中指出：所谓电子图书馆是一个尽最大可能提供信息存取的并通过电子技术扩大和管理信息资源的机构。在许多场合下，数字图书馆和电子图书馆可以等同使用。但电子图书馆侧重对收藏特色的概括，收藏品基本为电子读物，阅读手段一般通过电脑等，不一定提供网上信息或上网服务。

2. 虚拟图书馆

虚拟图书馆 (Virtual Library) 是指图书馆服务不局限于本馆的物理意义的馆藏，而是通过通信网络连接各图书馆、各地区、全国乃至全球信息资源的逻辑意义上的馆藏，用户可以在其中检索到比本馆馆藏多得多的信息。有人认为“虚拟图书馆”的概念最早是 1990 年提出来的，也有人认为是 1980 年不列颠图书馆外借部计算机与数据通讯工作负责人 A. J. Harley 提出来的。

3. 无墙图书馆

无墙图书馆 (Library without walls) 的概念主要从用户感觉的角度来刻画数字图书馆的部分特征，即网络化特征。利用 Internet 打破了获取信息的地理障碍和时间限制，使用户觉得好像正在利用一个没有围墙、不规定借阅时间的图书馆。可见，无墙图书馆反映了数字图书馆的部分特征，并不能代表数字图书馆的全部特征，所以，把数字图书馆完全等同于无墙图书馆是片面

的，不妥当的。

4. 虚拟现实图书馆

如果说数字图书馆与虚拟图书馆在含义上还有部分相似的话，那么，虚拟现实图书馆则是完全不同的概念。虚拟现实图书馆（Virtual Reality Library）是指基于虚拟现实技术的系统。它是为了模拟人们浏览书架的习惯，运用虚拟现实技术建立起来的。它应该是能够通过一种界面，存取某个典型图书馆馆藏的书目数据，这个显示在屏幕上的界面，看起来好像排满书架的书库，通过使用类似三维鼠标的某种装置，穿行于书架之间，浏览和阅读书中的内容，并加以控制。显然，虚拟现实图书馆完全不同于虚拟图书馆。

1.1.4 数字图书馆的基本特点

传统图书馆一般具有固定的馆舍，馆藏以印刷型书刊资料为主，服务以图书馆为中心，读者要借阅书刊、查询资料，需要到图书馆来。数字图书馆与传统图书馆的关系是既有区别又有联系。

传统图书馆收集、存储并重新组织信息，使读者能方便地查到它所想要的信息。传统图书馆同时跟踪读者使用情况，以保护信息提供者的权益。数字图书馆收集或创建数字化馆藏，这集成了各种数字化技术，如高分辨率数字扫描和色彩校正、光学字符识别（OCR）、信息压缩、转化等；建立在关系数据库系统上的数字信息组织、管理、查询技术能够帮助用户方便快捷地查找信息，并将信息按照用户期望的格式发送；在安全保护、访问许可和记帐服务等完善的权限管理之下，经授权的信息利用 Internet 的发布技术，实现全球信息共享。数字图书馆与图书馆自动化概念不同。图书馆自动化强调图书馆业务管理的计算机化，而数字

图书馆侧重于图书馆信息资源的数字化,突出信息的计算机存储处理、网络传输等特点。从某种意义上说,图书馆自动化只是数字图书馆实现的前提。数字图书馆与传统图书馆、自动化图书馆的比较如表 1-1 所示。

表 1-1 数字图书馆与传统图书馆、自动化图书馆的比较

	传统图书馆	自动化图书馆	数字图书馆
工作中心	馆藏	馆藏	用户
馆藏形式	印刷型	印刷型及少量电子出版物	数字信息资源
工作方式	手工作业	对书目数据及专题数据库进行自动化加工	对文献内容进行自动化加工
检索手段	手工检索卡片	对书目数据及专题数据库进行自动化检索	对文献内容进行智能检索
服务对象	为到馆读者服务	以到馆读者服务为主,在一定范围内提供文献传递服务	面向全球读者提供网上服务
馆藏加工	不加工	基本不加工	加工,并使馆藏具有增值效应

与传统图书馆相比,数字图书馆具有如下特点:

1. 信息存储方式由纸张上的文字变成了磁性介质上的电磁信号,这为压缩存储空间、改进组织方式、提高检索速度、方便用户远程检索、加快更新维护、降低维护费用等提供了基础。

2. 信息组织形式从纸张上顺序的线性方式转变为计算机的网状结构方式。索引文件、超文本技术等,使得信息可以按本身的逻辑关系组织成相互联系的网状结构,这为方便用户检索、提高检索效率奠定了基础。

3. 馆藏不仅包括纸质的非数字化的文献,还将扩展到不能以印制品形式纪录和传播的数字化作品,如:文本、图形、图像、声音、视频、商品目录等多种媒体信息。多种信息可以通过