

# 数字图书馆的建设与应用

主 编 李 辉

副主编 王 璐 李英珍 夏 波  
黄 琦

参 编 (以姓氏笔画为序)

王雪梅 史少华 李振东

刘 燕 师宝玉 杨利超

郑文芬 贾淑敏 梅雪晴

温 洁 谢朝晖 薛 瑞

沈阳出版社

## 前 言

人类社会的 20 世纪的最后十年发生了深刻的变化,正步入以“网络”、“数字”为标志的信息社会。随着当代信息技术的发展,计算机网络正在向高速化、多媒体化、以宽带主干网为基础的多网融合的方向发展,将逐步实现 5 个 W 的个人通信:即任何人(whoever)可以在任何时间(whenever)、任何地点(wherever)和任何一个其他地方(homever)通过网络进行通信,以传递任何信息(whatever)。

随着计算机技术的迅猛发展,特别是网络技术、数码存储与传输技术等的全 面普及,使得人们对文献信息的加工、存储、查询、利用等方面有了新的要求,数字图书馆(Digital Library DL)应运而生。信息资源共享,已成为 21 世纪人类社会发展的必然趋势,也是人类社会长期追求的目标。数字图书馆建设就是通过计算机网络化结构形式,将不同地区、不同国家的图书馆连接起来,形成一个跨区域、跨国家的网络。人们可以在网络上利用一切信息资源,不受时间和空间的限制,因而实现真正意义上的资源共享。

首先,推动数字图书馆研究的最主要动力是 Internet 的发展。网络的互联使访问分散在各处的信息资源成为可能,但这些各自独立的信息存储具有各自不同的组织、描述和检索方式,所收藏信息的质量也良莠不齐。网络环境下跨库存储的、统一的、高效的访问和利用工具,以及高质量信息的生成、组织和提取成为研究的重点,而这正是数字图书馆的研究内容。如果把 Internet 看成是一个巨大的无墙图书馆,广义的数字图书馆的目标就是要优化 Internet 的信息存储结构,提供一致的检索接口,使整个网络成为一个虚拟的、单一的、有组织、有结构的信息集合,实现跨存储的无线查找。

数字化革命和因特网的大发展,带来了经济、贸易、信息传播的全球化,深刻影响着社会的各个层面。巨量信息涌入因特网,使其信息极为丰富,也使信息查询、检索变得十分困难,尤其是对于上网漫游的新手更是如此。这主要是由于因特网上的信息资源很大程度上处于混乱无序的状况,大大影响了信息利用效率。这一重大弊端,清楚地表明了因特网这种全球信息服务网络,急需一个组织管理

信息资源的新模式,上网的资源应按这个模式组织管理。

其次,随着现代信息技术革命的深入发展,图书馆面临巨大的挑战。以书刊资料为主要收藏载体的传统图书馆,逐渐难以适应数字时代的要求,需要不断调整自身的馆藏结构和服务方式。图书馆正在发生前所未有的变革,正在向数字图书馆迈进。

数字图书馆是一项崭新的事物,它的提出,并不源自于图书馆自身,而是信息化社会不断发展所造成的,它的出现给图书馆事业的发展带来新的契机。尽管对于数字图书馆的认识还很不一致,但数字化对图书馆的影响是无可回避的事实,图书馆的数字化是一个趋势,必将成为21世纪图书馆事业发展的主旋律。

也就是说,一方面是满足当代信息技术高速发展、知识经济初见端倪的迫切要求,同时也是图书馆界顺应时代发展、积极拓展自身生存空间的需要。这就是提出数字图书馆的主要背景。

本书在编写过程中由于时间、水平有限,书中难免有错误和疏漏的地方,望各位专家、读者给予指正。在编写过程中参考了大量专家的资料,在此表示衷心的感谢。

本书由李辉进行统稿和定稿,由各位老师负责纠错改正。本书具体分工如下:

- 李 辉(河南科技学院图书馆),第十九章、第二十章;
- 王雪梅(南阳医学高等专科学校),第十四章第三节、第五章第五节;
- 王 璐(河南职业技术学院图书馆),第二十四章、第二十五章第二节;
- 史少华(河南第一荣康医院图书室),第二十五章第一节;
- 李英珍(河南师范大学图书馆),第二章、第三章;
- 李振东(新乡学院图书馆),第二十一章;
- 刘 燕(安阳市图书馆),第一章、第十二章;
- 师宝玉(河南省图书馆),第十六章、第十七章;
- 杨利超(河南省图书馆),第七章;
- 郑文芬(社旗县图书馆),第八章;
- 夏 波(大连职业技术学院图书馆),第十八章、第二十三章;
- 贾淑敏(河南中医学院图书馆),第十一章、第十五章;
- 黄 琦(南阳医学高等专科学校),第四章、第五章第一、二、三、四、六节;
- 梅雪晴(周口川汇区图书馆),第六章、第十四章第一、二节;
- 温 洁(中州大学图书馆),第十章、第十四章第四节;
- 谢朝晖(河南平顶山学院图书馆),第九章、第十三章;
- 薛 瑞(周口中医院图书馆),第十四章第五节、第二十二章。

# 目 录

|                           |    |
|---------------------------|----|
| 第一章 数字图书馆及其相关概念 .....     | 1  |
| 第一节 数字图书馆产生的背景 .....      | 1  |
| 第二节 数字图书馆定义及相关概念 .....    | 5  |
| 第三节 数字图书馆与传统图书馆的比较 .....  | 11 |
| 第二章 数字图书馆技术 .....         | 16 |
| 第一节 概述 .....              | 16 |
| 第二节 基于信息网格的数字图书馆 .....    | 21 |
| 第三章 数字图书馆的建设探索 .....      | 34 |
| 第一节 数字图书馆的解决方案概述 .....    | 34 |
| 第二节 几种数字图书馆的解决方案介绍 .....  | 35 |
| 第三节 数字图书馆的资源建设 .....      | 37 |
| 第四节 我国数字图书馆的建设情况 .....    | 42 |
| 第五节 数字图书馆建设面临的技术挑战 .....  | 53 |
| 第四章 数字图书馆的信息资源建设 .....    | 59 |
| 第一节 数字图书馆的信息资源内涵 .....    | 59 |
| 第二节 信息资源建设的理论与模式 .....    | 64 |
| 第三节 数字信息制作、检索与保存 .....    | 66 |
| 第五章 数字图书馆网络系统建设 .....     | 74 |
| 第一节 数字图书馆网络系统建设研究 .....   | 74 |
| 第二节 数字图书馆的关键技术研究 .....    | 78 |
| 第三节 数字图书馆智能化建设的项目需求 ..... | 80 |
| 第四节 数字信息归档系统 .....        | 82 |
| 第五节 多媒体高速数据网络体系 .....     | 84 |
| 第六节 数字图书馆用户界面研究 .....     | 86 |

|                                  |     |
|----------------------------------|-----|
| 第六章 数字图书馆的网络信息安全 .....           | 94  |
| 第一节 网络信息安全现状与防范机制 .....          | 94  |
| 第二节 网络信息安全功能和安全分析 .....          | 101 |
| 第七章 数字图书馆与传统图书馆的关系 .....         | 109 |
| 第一节 数字图书馆与传统图书馆的区别与比较 .....      | 109 |
| 第二节 数字图书馆对信息服务的影响 .....          | 112 |
| 第三节 数字图书馆是否能完全取代传统图书馆 .....      | 114 |
| 第八章 馆藏建设的原则与构成 .....             | 117 |
| 第一节 馆藏建设的基本原则 .....              | 117 |
| 第二节 馆藏信息资源的整合 .....              | 119 |
| 第三节 馆藏信息资源结构体系 .....             | 122 |
| 第九章 数字图书馆的建设规范 .....             | 126 |
| 第一节 数字图书馆标准规范的建设 .....           | 126 |
| 第二节 我国数字图书馆标准规范结构体系 .....        | 140 |
| 第三节 数字对象的标准规范 .....              | 149 |
| 第十章 数字图书馆的建设探索 .....             | 162 |
| 第一节 数字图书馆的解决方案概述 .....           | 162 |
| 第二节 数字图书馆的资源建设 .....             | 168 |
| 第三节 我国数字图书馆的建设情况 .....           | 172 |
| 第四节 数字图书馆建设面临的技术挑战 .....         | 176 |
| 第十一章 数字馆藏资源的存储与组织 .....          | 179 |
| 第一节 存储体的功能 .....                 | 179 |
| 第二节 电子书库构建与网络存储 .....            | 184 |
| 第三节 信息资源的组织管理 .....              | 190 |
| 第十二章 数字图书馆的研究与应用实例 .....         | 202 |
| 第一节 数字图书馆研究 .....                | 202 |
| 第二节 数字图书馆应用实例 .....              | 208 |
| 第三节 数字图书馆企业化运营概述 .....           | 213 |
| 第十三章 企业化运营是数字图书馆的正确选择 .....      | 220 |
| 第一节 如何理解数字图书馆企业化运营 .....         | 220 |
| 第二节 如何实现数字图书馆企业化运营 .....         | 225 |
| 第三节 数字图书馆企业化运营面临的主要问题及应对策略 ..... | 232 |
| 第十四章 数字图书馆的版权保护 .....            | 238 |
| 第一节 数字图书馆馆藏资源建设的知识产权 .....       | 238 |
| 第二节 数据库建设版权问题 .....              | 243 |

|      |                       |     |
|------|-----------------------|-----|
| 第三节  | 数字图书馆用户服务的知识产权 .....  | 256 |
| 第四节  | 数字图书馆版权保护措施 .....     | 261 |
| 第五节  | 数字图书馆建设中知识产权案例 .....  | 278 |
| 第十五章 | 数字图书馆运维与评估 .....      | 284 |
| 第一节  | 数字图书馆的系统构成 .....      | 284 |
| 第二节  | 数字图书馆的运用 .....        | 288 |
| 第三节  | 数字图书馆的评估 .....        | 290 |
| 第十六章 | 数字图书馆数字资源构建 .....     | 295 |
| 第一节  | 数字资源的特点及类型 .....      | 295 |
| 第二节  | 数字资源馆藏建设规划 .....      | 297 |
| 第三节  | 数字资源的采购 .....         | 300 |
| 第四节  | 数字资源建设中应注意的问题 .....   | 302 |
| 第十七章 | 数字图书馆服务功能设置 .....     | 310 |
| 第一节  | 数字图书馆服务特点 .....       | 310 |
| 第二节  | 数字图书馆资源导航服务 .....     | 315 |
| 第三节  | 数字图书馆检索服务 .....       | 317 |
| 第四节  | 数字图书馆主动推送服务 .....     | 320 |
| 第五节  | 数字图书馆个性化信息服务 .....    | 322 |
| 第六节  | 图书馆 2.0 下的新型服务 .....  | 326 |
| 第十八章 | 数字图书馆的标准和规范 .....     | 328 |
| 第一节  | 数字资源唯一标识符的标准和规范 ..... | 328 |
| 第二节  | 专门数字对象描述元数据 .....     | 332 |
| 第三节  | 管理元数据规范 .....         | 337 |
| 第四节  | 对象数据建设的标准与规范 .....    | 340 |
| 第五节  | 元数据互操作、传输协议 .....     | 342 |
| 第六节  | 电子文件管理元数据规范 .....     | 345 |
| 第十九章 | 数字馆藏信息资源评价体系 .....    | 357 |
| 第一节  | 网络环境下馆藏评价内容与沿革 .....  | 357 |
| 第二节  | 数字馆藏的质量评价标准 .....     | 361 |
| 第三节  | 网络信息资源评价的理论与实践 .....  | 362 |
| 第四节  | 数字图书馆馆藏评价方法 .....     | 369 |
| 第五节  | 网络信息资源的安全管理 .....     | 376 |
| 第二十章 | 数字图书馆的信息资源 .....      | 383 |
| 第一节  | 图书馆信息资源的含义 .....      | 383 |
| 第二节  | 图书馆信息资源 .....         | 386 |

|       |                            |     |
|-------|----------------------------|-----|
| 第三节   | 信息资源的分类 .....              | 388 |
| 第四节   | 数字图书馆信息资源的分析 .....         | 390 |
| 第二十一章 | 数字图书馆的版权责任 .....           | 397 |
| 第一节   | 案例链接 .....                 | 397 |
| 第二节   | 数字图书馆版权责任的归责原则 .....       | 398 |
| 第三节   | 充当不同角色的数字图书馆的版权责任 .....    | 399 |
| 第四节   | 《解释》数字图书馆版权责任 .....        | 400 |
| 第五节   | DMCA 与数字图书馆版权责任 .....      | 402 |
| 第二十二章 | 数字图书馆版权保护对策 .....          | 404 |
| 第一节   | 图书馆对作品数字化利用中存在的主要问题 .....  | 404 |
| 第二节   | 图书馆立法 .....                | 414 |
| 第三节   | 版权制度的调整 .....              | 420 |
| 第四节   | 提高图书馆员的版权素质 .....          | 424 |
| 第五节   | 正确运用法律和版权风险评估 .....        | 425 |
| 第六节   | 版权管理制度和读者控制 .....          | 426 |
| 第七节   | 图书馆版权政策和版权认证 .....         | 428 |
| 第八节   | 图书馆版权联盟 .....              | 428 |
| 第九节   | 版权保护的国际合作 .....            | 429 |
| 第二十三章 | 数字图书馆建设和隐私权保护 .....        | 431 |
| 第一节   | 个人数据收集中的隐私权问题 .....        | 431 |
| 第二节   | 个人数据传输和贮存中的隐私权问题 .....     | 432 |
| 第三节   | 个人数据利用中的隐私权问题及对策 .....     | 432 |
| 第二十四章 | 数字图书馆网络化建设 .....           | 436 |
| 第一节   | 智能化图书馆的基础——综合布线系统的建设 ..... | 436 |
| 第二节   | 光纤的性能指标 .....              | 438 |
| 第三节   | SAN 的概述 .....              | 452 |
| 第二十五章 | 数字化图书馆服务平台的搭建 .....        | 463 |
| 第一节   | 高校图书馆电子阅览室的概况 .....        | 463 |
| 第二节   | 手机服务的图书馆中的应用 .....         | 474 |
| 参考文献  | .....                      | 480 |

# 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 数字图书馆产生的背景

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。

### （一）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,加世纪 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”“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,在德国的艾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他一些著名学会、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)等几个大学术组织在 Bethesda 召开了规模很大的首届 ACM 数字图书馆国际会议。

## (二) 技术背景

20 世纪 80 年代以来,光电技术开始得到广泛应用,光钎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的应用。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,可以由那些被即时的搜集、交流并提交无需存储的数据(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商的信息链接到另一个服务上的信息。随着 Inter-

net 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 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、图像三种不太能够的媒体结合起来,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种重要方面,它使图书馆馆藏结构产生了重大变化。

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,数据库愈发显示出它的重要性。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。各个系统能够的数据库通过卫星、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。

### (三)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。纵观近代图书馆的发展历程,我们可以放哪关照托夫勒的说法,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人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,第一阶段可成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,也成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 年扩大为六哥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,为数字图书馆技术向实用化做了准备。

20 世纪 60 年底啊电子书刊的兴起,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,当时就提出来了“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 Tonpson)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现实》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、霍国庆所著的《现代图书馆理论》,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。

#### (四) 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,仙子啊书目已达 3000 万条之多。

近年来,随着计算机技术、网络技术的飞速发展,有其是 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。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,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)建设工程业已启动,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 数字图书馆定义及相关概念

### 一、数字图书馆的定义

数字图书馆(Digital Library,简称 DL)是当前最热门的话题之一,相关的定义有很多种。然而,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。自 1990 年以来,有近百种的看法,较有代表性的观点有以下几个:

1990 年美国密执安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)存在全球网络上的、以多媒体数字化格式存在的、为数巨大的,且仍在不断增多的信息。

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作了如下定义:“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。”一年后此定义又修改为:数字图书馆是一种多媒体数据与信息管理方法的有序结合,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。

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。

俄罗斯多数学者认为,数字图书馆是一个分布式信息系统,它能够保存并有效地利用各种各样的电子文献,这些电子文献可以使最终用户通过全球网络传输便利地获取。

日本图书馆情报大学的田佃孝一认为数字个是指不局限于文本,而且包含画像、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集合进行处理的图书馆。

我国前文化部副部长徐文伯认为,数字图书馆是以数字形式存储和处理信息的图书馆,即对有高度价值的图像、文本、语言、音响、影视、软件和科学数据等多媒体信息进行收集,组织规范性加工,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、数字安全管理范畴。

同时,基于不同的出发点,理论界对数字图书馆的定义归纳起来有以下几种:

第一种定义:数字图书馆的概念是组织数字化信息及其技术进入图书馆并提供有效服务,几乎图书馆的所有载体信息均能以数字化的形式获得,读者通过网络组织访问外界数字化图书馆和文献信息数据库系统。图书馆和网络系统连接,让人们很方便地共享资源。

第二种定义:数字图书馆是一个大系统,它是分布式的、大规模的和有组织的数据库和知识库,用户和用户团体可以对系统内的数据库进行一致性访问,获得自己所需要的最终信息。

第三种定义:数字图书馆强调它存储的信息是用电磁介质,按二进制编码的方法加以存储和管理,是把原先用纸张存储的信息转变为用计算机和多媒体技术存储和处理的信息。

第四种定义:数字图书馆是一个数字化的信息系统,它将分散于各种载体、不同地理位置的信息资源以数字化方式储存、以网络化方式相互连接,提供即时利用,实现资源共享。

第五种定义:利用当今先进的数字化技术,通过诸如国家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大量的分散在不同储存处的电子物品的全部内容。

由于数字图书馆具有广泛的内容和涵义,因而现在对数字图书馆有各种各样的定义。综观各家之言,结合数字图书馆的特征与结构,我们可以给数字图书馆下这样一个定义:数字图书馆是传统图书馆功能的扩展,它对信息进行收集、转换、描述,并以计算机可处理的数字化形式存贮馆藏信息和网络数字化信息,以智能化的信息检索方式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服务。

## 二、数字图书馆的特征

数字图书馆虽然称为图书馆,但已不同于传统意义上的图书馆,它通过利用新技术,以新的方式执行图书馆的功能。数字图书馆的提出和实际建设是与计算机技术、通信和网络技术、高密度存储技术和多媒体技术的飞速发展密切相关的,它明显地带有新技术革命的时代烙印。因而具有如下特点:

### (一)传递网络化

数字图书馆不是一个单独的物理实体,而是一个信息空间、互联空间,这是网络技术赋予它的特性,使得用户对信息的查询不受时空的限制。

网络和通讯系统是数字图书馆的重要基础,若没有网络系统或网络系统不完善,数字图书馆就失去了它的真正意义。信息的网络传输使数字图书馆超越时空观念,跨越馆藏信息的地域界限。通过网络传输,使用户必须到图书馆查阅信息,变成了读者可以在世界上的任

何地方、任何时间查阅任何一个开放的数字图书馆的信息,大大缩短了信息传递时间以及信息提供者和使用者的距离,从而加快了信息交流与反馈的速度。通过地区、国家和国际网络,数字图书馆和各个信息服务中心以及分散的信息数据库系统连接起来,方便地收集其他机构的信息,充实自己的馆藏,提供信息的增值服务。

## (二)信息资源数字化

数字图书馆的信息是经过计算机处理的文字、数值、色和彩色图形、静止和活动图像、声音、电影、动画,还包括现有纸介质、缩微、平片等经数字化处理的信息,它们都以二进制(0/1)形式编码、存储,并经组加工放在磁或光存储器上,需要时可以用多种便捷的方式提供给读者。这是数字书馆的基本特征。

## (三)直接、网状的信息组织方式

信息的组织形式从原先纸张的顺序、线性方式转变为直接的、网状的组织方式。电子计算机的存储为信息的直接存取提供了可能性能性,从排列文件、直接存取文件、数据库到超文本技术,使信息可以按本身的逻辑关系组织成相且联系的网状结构。

## (四)服务模式以用户为中心

开发数字图书馆的白的是使用户高度灵活地获取利使用信息,因此数字图书馆的服务模式足以用户为中心的。用户通过数字图书馆提供的广泛的用户接口,良好的检索功能,直观易用的人机界面,随时随地访问数字图书馆根据用户的需要组织的逻辑的、动态的信息资源。信息资源现有数字图书馆本身的馆藏,也百数字图书馆通过网络获得的非本馆的信息。这使得数字图书馆的服务突破了本馆的局限。当用户在联机查找遇到问题时,馆员能进行干预,为用户解决问题。

# 三、数字图书馆的功能及作用

## (一)数字图书馆的功能

### 1. 数字式资源的采集

目前数字图书馆的馆藏来源有两种途径,一是把大量现已存在的以不同形和载体存储的信息资料,如纸制品资料、图片、声像资料等转化成数字形式存储起来。直接用于该任务的技术包括文字图像扫描处理、图像和语言的识别以及对数字化初始信息的各种再加工技术。在 IBM 数字化图书馆中把这一功能称为获取;二是提供直接生成的数字化信息以便计算机直接使用,如购买电子出版物,从络上获取信息,图书馆自动化建设中直接形成的 MARC 格式的馆藏目录数据库一些专题数据库等。

### 2. 数字式资源的存储与管理

与传统图书馆不同,数字图书馆的信息资源足存储在计算机、磁带、磁盘光盘中,信息存储量是以存储多少 GB 字节来计算的对数字图书馆来说,其存储量应该是几百个、几千个 GB 甚至更大,即海量信息。随着信息量的巨增,存储规模也越来越大,已用万亿(或太)字节来表示。存储这些海量信息主要涉及存储设备的容量,硬件随机读取的速度以及信息的组织与管理方法等存储技术。当前,数字图书馆大多数采用客户机/服务器模式,客户、图书馆服务器和对象服务器构成信息传递的核心结构。图书馆服务器主要管理数据的目录、索引和查询,而对象服务器用于管理数字化的对象。客户向图书馆服务器提出要求,图书馆服务器

将客户申请转发给对象服务器,最后由对象服务器将检索到的数据送发给客户机。这种模式的好处在于:随着信息量的增加,存储规模的扩大,在需要时随时随地可增加对象服务器和客户机,而整个系统的工作并不受影响。海量信息的存储方法是最为关注的问题,在存储这些数字化信息时,要考虑到计算机硬件和软件应具有广泛的接口来保持它们的有效性,因为数字信息是依靠硬件和软件来使用它们的,还要考虑到检索信息时的响应速度,应该具有接使用频率调动存储位置的能力。全文存储是目前数字图书馆存储的主要方式它包括文本信息和图像信息混合的电子全文信息,以建立 SGML 等电子文本格式基础上的全文存储为重点。

### 3. 数字式资源的访问和查询

有效的文本数据库查询和多媒体资料的查询策略是数字图书馆的重要技术。由于信息的多样性,原有的文本类型的文件检索,传统的基于关键字索引以及通过布尔表达式的查询方式已无法满足用户的查询要求,现在正向全文检索方向发展。新型的全文检索主要有三种实现方法:自由文本查询,即用自己指定的关键字字符串直接与全文文本的一次数据高速对照进行检索,它的查询结果是一个按序列值排列的文件列表;对文本内容中的每个词进行位置扫描,然后排序,最后建立以每个词(字)的离散码为表目的倒排文件;采用基于 HTML 语言的超文本模型建立的全文数据库,使用户便捷地查看到查询结果。为多媒体信息建立索引,便于用户快速和简便地对声音、图像等信息内容进行筛选和确定,这是当前研究的重点问题之一。直接针对图像、声音建立索引,可以根据颜色形状、纹理等在图像中的位置对图像进行查找,例如用颜色分布来索引纺织品的图案;利用平均色特性来查询图像中的主导色;按音乐的曲调旋律进行查询问题等。用同样的方法对数字化的音频、视频信息内容进行查找,是访问数字图书馆的关键技术。

### 4. 数字式资源的传送与发布

数字图书馆的信息资源是通过网络系统传送到用户面前,从而实现信息的发布。当前的 Internet 就是数字图书馆现实的网络环境,它是一大批共同遵守 TCP/IP 协议的计算机网络互联而成的庞大网络。今后的网络和通讯环境是各国的 NI(国家信息基础)和全球的 GI(全球信息基础)。数字图书馆的传送主要涉及到网络技术的应用,重点关注对图像、音频和视频等多种媒体信息的传输。多媒体信息传送要求较大的带宽,如 CD 需要 14MBPS 带宽, MPEG 需要 12—40MBPS 带宽。在实际中,至少单向 14MBPS 的音频、视频总带宽进行传输,才能看到质量较好的视频图像。实时音频和视频传送不仅对网络的速度有很高的要求,同时对提供信息的相关服务器的共享性也有很高的要求,如 IBM 数字图书馆支持各类平台的多通讯技术,支持数字及模拟信息传送的 ATM 技术及网络管理软件,具有扩充先进网络的能力。目前 ISDN 综合服务数字网已成为多媒体通讯的基本传输网络,现在正向 B-ISDN,使网络的传输和交换率达到每秒百兆比特以上。

### 5. 数字式资源的权限管理

由于数字图书馆的主要结构原则是对不同类型开放环境的操作能力,为用户共享信息资源,获得信息服务提供了极大的方便。因此对数字图书馆来说,安全性显得非常重要。数字图书馆具有一般计算机网络系统的管理功能,旨在对访问权的控制,该问题在利用原有计算机读取管理技术以及域名管理技术的基础上已得到了较好的解决。更重要的是对知识产权的保护,通过采用技术手段来防止未经授权使用版权人的资源,用来保护信息拥有者和最

终用户的利益。

## (二)数字图书馆的体系结构

### 1. 数据存储与管理

数字图书馆在数据的存储和管理上,采用目前已被业界普遍接受的“浏览器/web 服务器/数据库服务器”模式结构。读者客户机、图书馆 Web 服务器和数字信息数据库服务器构成信息传递的核心结构。Web 服务器主要接受读者客户端的查询请求、进行数据处理和处理结果的发送,管理 HTML,构成的信息空间。提供对数据库的存取接口;数据库服务器主要负责管理数字化馆藏,包括全文数据、多媒体数据、标准的书目数据、二次文献数据、事实数据等。它通过接收 Web 服务器的请求,对数据进行处理,然后把处理结果传送给 Web 服务器;读者客户端通过各种网络实现与 Web 服务器的连接。通过浏览器访问 Web 服务器提供的各种功能和丰富的数字化馆藏。数字图书馆 Web 服务器与数据库连接的存取接口,可采用 CGI 公共接口、通过 API 和 ODBC 访问数据库、利用 Java 机制设计的标准 SQL 数据库连接接口 IDUC 访问数据库等 3 种各具特色的途径实现。

### 2. 应用系统

数字图书馆具有庞大的数据量,要对巨量数据进行存储、管理和索引,并为众多的网上读者提供快速的信息检索服务,只用一套系统是不可能完成的,而应综合多个应用系统的功能。众多研究表明,数字图书馆应建立“一束”应用系统,以满足众多读者的各种需要。应用系统束一般由以下几部分构成:创建数据,将非数字化信息转换为数字化,如文本录入、图像扫描、音频视频数字化等;数据描述,使用一定的语言描述创建的数字化信息,目前大多采用 GML 描述语言;全文检索和多媒体信息检索,支持 SGML、PDF、XMI、HTML、ASCII 等格式的全文检索和直接针对图像、声音建立索引,按照颜色、形状、纹理在图像中的位置对图像进行检索等;系统记帐和维护管理等。

### 3. 馆藏结构

数字图书馆馆藏建设是数字图书馆建设的再要内容。随着电子出版物的大量涌现和 Internet 网络的飞速发展,文献出版类型呈现多样化趋势,按其载体形式主要可划分为印刷版、光盘版和网上电子版两大类。数字图书馆的馆藏主要由经过数字化转换的现有文献数字馆藏、本来就电子形式出版的光盘等形式的电子出版物和网上电子出版物、新出版的文献或经过回溯性加工的资料等几个部分组成,其资源包括期刊、专著、参考工具书、音频、视频等类型。

### 4. 组织机构

为保证数字图书馆建设的顺利进行,必须抽调专门的人员组成专门的机构,该机构按工作流程和任务可设立以下 4 个部门:

数据采集部。主要负责收集整理馆藏待数字化转换信息、各类型电子出版物和 Internet 网上的相关信息等。

数据处理部负责将采集到的信息进行转换、描述和加工处理,包括将现有文献信息通过扫描仪数码相机等设备转换存贮于计算机;审校已转换的数据;分析文献中的有效信息,抽取文本并用 SGML,或其他描述语言对析出内容进行描述。

信息服务部。负责数字图书馆服务项目的设计;制作、维护和更新图书馆的 web 网页;开展读者培训,辅导读者检索,收集读者反馈意见和建议等。技术研究部。负责数字图书馆

新技术、新工具、新标准的研究与追踪;系统设备的更新与维护;数据处理系统的升级与维护;数字化信息的保存与保护。

### (三) 数字图书馆的相关标准及其工具

从某种程度上来说,数字图书馆是建立在因特网基础之上的。因特网是世界上最大的信息宝库,是对数据组织得较好的一种系统,但它毕竟是淹没在全球的信息资源的海洋中。如何快捷、准确地从信息资源中找到自己最需要的信息,已成为许多专家、学者、公司、用户共同关心的热点。

在数字图书馆中或任何 Web 站点,信息组织必须按照某种格式标准,这就产生了对文献描述和处理语言的国际标准问题。像当前研究的热点之一“Metadata”,它是描述 Internet 上的数据和资源,促进 Internet 信息资源的组织和发展的数据;搜索引擎将是数字图书馆的重要工具;在数字图书馆中,运用全文检索系统、超文本、超媒体技术,将提高读者检索的查全率和查准率;另外,利用统一的界面来查找网上各种数字图书馆的资源也是人们关心的问题。

#### 1. 数字化文献描述和处理语言

数字化文献描述和处理语言主要有:一、标准通用置标语言 SGML 及其相关标准,它包括许多 ISO 的标准,HTML 是它的子集,专门用来为 WWW 建立超媒体文件的语言。二、扩展的置标语言 XML,它由 W3C 组织提出并用来定义处理 Metadata 的基本原则和补充。三、资源描述框架(RDF)模型与句法,它是在 W3C 的授权下,制定出的并将成为最终的 Metadata 规范。四、信息检索。应用服务定义和提议技术规范(ANSI/NISO Z39.50:ISO23950),它是一个为适应在网络上不同系统互联、互访的协议。五、通用多八位编码字符集(UNICODE),它包含了汉字等多国文字,字符总共有 29092 个。

#### 2. Metadata:描述数据的数据

通过 Metadata:使我们能够了解到某个 Internet 站点的资源类型,某个 Web 页面的标题、作者、主题、关键词及内容摘要等等信息。分布在全球 Internet 上的 Web 页面,像在一个庞大图书馆有许许多多电子文献,需要有序的、按一定标准组织起来。有专家指出,Metadata 是关于数据的数据,需要图书馆去组织和发现 Internet 网上信息资源。目前有许多团体研究 Metadata,其中 Dublin Core 是一个国际范围内通用的适用于资源发现系统的 Metadata 标准。它由 15 个 Metadata 标记元素组成的系统,但它适用于 HTML 的环境,在此环境中还有如 PICS 标准。另外,随着 XML 应用环境的日渐成熟,许多基于 XML 的 Metadata 规范也被提出来。如 Web Collection, CDF 等。许多团体提出了不同的 Metadata 规范,虽有相似形,但彼此不能兼容。在 W3C 的授权下,一些 Metadata 的团体集思广益,制定出符合多种需要、又灵活的 RDF 结构,用以支持因特网和 WWW 上的 Metadata。

#### 3. 搜索引擎

搜索引擎是未来数字图书馆组织和发现网上电子文献信息资源的重要工具,特别是近年来出现的动态建立索引的搜索引擎,它能自动帮助数字图书馆组织和发现新资源。目前在因特网上有几十种,大致可分为四种搜索万维网信息的方式:浏览式;按主题指南分类目录进行查询;利用检索软件进行关键词或自然语言的查询;集成式、多线索的查询等。

#### 4. 全文检索系统

现代全文检索系统已引入超文本和超媒体的概念,它不但对本地数字图书馆内的文献