



数字信息资源的开发与利用研究

主 编 王云娣

2



WUHAN UNIVERSITY PRESS

武汉大学出版社

G203

16

数字信息资源的开发与利用研究

主 编 王云娣

编著者 王云娣
胡秀青
黄光明



WUHAN UNIVERSITY PRESS

武汉大学出版社

图书在版编目(CIP)数据

数字信息资源的开发与利用研究/王云娣主编. —武汉:武汉大学出版社, 2005. 8

ISBN 7-307-04654-7

I. 数… II. 王… III. 信息管理—研究 IV. G203

中国版本图书馆 CIP 数据核字(2005)第 078140 号

责任编辑:严红 陈鹏 责任校对:王建 版式设计:支笛

出版发行:武汉大学出版社 (430072 武昌 珞珈山)

(电子邮件:wdp4@whu.edu.cn 网址:www.wdp.com.cn)

印刷:湖北省新华印务有限责任公司

开本:850×1168 1/32 印张:12 字数:278千字 插页:1

版次:2005年8月第1版 2005年8月第1次印刷

ISBN 7-307-04654-7/G·749 定价:20.00元

版权所有,不得翻印;凡购买我社的图书,如有缺页、倒页、脱页等质量问题,请与当地图书销售部门联系调换。

内 容 提 要

本书系统地论述了数字信息资源的起源与发展、类型与特点;数字信息资源的组织;数字信息资源的检索技术与检索策略;电子图书与电子期刊的开发与利用;搜索引擎的利用与评价;数字文献共享与文献传递服务等方面内容。除此而外,对网络学科信息门户、隐性网站资源的开发与利用、数字信息资源的合理使用及网络开放存取资源的利用等现实问题,也进行了分析和阐述。

本书除了适用于高等学校图书馆学、情报学等相关专业作为教科书和参考书外,还非常适合作为信息素质教育的基础理论课教材,亦可供图书情报机构及其他信息服务从业人员学习使用。

前 言

随着全球信息化进程的发展,人们越来越深刻地认识到,信息作为与材料和能源同等重要的战略资源,不仅是社会的重要财富和资产,而且是最活跃的生产力要素之一。信息资源对社会发展的作用日益突出,已成为新的开放环境下政治、经济、文化和军事等国际竞争的焦点。

近年来,电子出版、网络出版的飞速发展和文献数字化步伐的加快,促进了全球数字信息资源的急剧增长。美国加利福尼亚大学伯克利分校信息管理及系统学院 Peter Lyman 和 Hal R. Varian 领导的研究人员发现,2002 年,全球由纸张、胶片、磁和光存储介质所记录的信息总量达到 5 万亿兆字节,约等于 1999 年全球信息产量的两倍,其中 92% 的信息资源存储在磁介质上(主要是硬盘上),数字信息资源所占的比重越来越大。数字信息资源的飞速增长,频繁更新,为人们全面和及时获取信息提供了有力的保障。数字信息资源已成为人们学习、工作、生活和科学研究的有力工具。

信息只有被利用才能体现其价值,然而,数字信息资源无序、易更改、不安全、缺乏规范和质量控制等特点,给信息资源的检索和利用带来极大的困难。因此,加强数字信息资源的组织,促进信息资源的开发与利用具有非常重要的现实意义。本书的写作目的就是希望通过对数字信息资源的种类、来源与专题进

行系统的分析,从技术方法上进行阐述并用例证说明,以便为广大用户在利用数字信息资源时提供有益的帮助。

为了有利于数字信息资源的开发和利用研究,笔者在跟踪国内外数字信息资源最新研究成果的基础上,结合自身的工作实践,确定了本书编写的指导思想和框架设计,并邀请胡秀青副研究馆员及黄光明副研究馆员,共同完成了本书的编写工作。具体分工如下:

王云娣负责编写第一章、第六章、第七章、第八章和第十一章;胡秀青负责编写第三章、第四章、第五章和第九章;黄光明负责编写第二章和第十章。全书由王云娣修改定稿。

初稿完成之后,武汉大学信息管理学院院长兼博士生导师陈传夫教授对该书提出了宝贵的修改意见,在此表示衷心的感谢!感谢武汉大学管理学院博士生导师詹德优教授的关心和鼓励。同时,本书的出版得到了浙江师范大学图书馆领导和同事的支持及帮助。感谢金明生研究馆员、吴奕宽副研究馆员为本书的编写提供的帮助,并对所有引用文献及参考文献的作者们表示谢意。武汉大学出版社严红副编审为本书的出版付出了辛勤劳动,在此谨表谢忱。

由于时间仓促,加上作者水平的限制和数字信息资源领域发展迅速,书中必有不少纰漏之处,作者诚恳地希望专家、学者和同行给予指正。

王云娣

2005年5月

目 录

前 言	1
第 1 章 数字信息资源概述	1
1.1 数字信息资源简介	1
1.2 数字信息资源的类型和特点	9
1.3 数字信息资源开发利用的必要性及利用 能力的培养	18
1.4 我国数字信息资源开发利用的现状 & 对策	29
第 2 章 数字信息资源的组织	40
2.1 信息组织的主要方式	41
2.2 数字信息资源的揭示	44
2.3 数字信息资源的描述——元数据	48
2.4 元数据在数字信息资源组织中的应用	62
2.5 数字信息资源的组织技术	68
第 3 章 数字信息资源检索技术与策略	74
3.1 数字信息资源检索概述	74
3.2 检索语言	80
3.3 数字信息资源检索技术	85
3.4 数字信息资源检索系统的检索功能	92

3.5	数字信息资源检索的策略	96
第4章	电子图书的开发与利用	106
4.1	电子图书概述	106
4.2	电子图书的出版格式及常用阅读软件	116
4.3	电子图书的运营模式	122
4.4	网上书店及电子图书系统	127
4.5	网上免费电子图书的利用	136
第5章	电子期刊的开发与利用	142
5.1	电子期刊概述	142
5.2	网上电子期刊资源与服务	150
5.3	电子期刊的检索	157
5.4	国外著名的英文电子期刊	162
5.5	我国电子期刊的发展现状	170
第6章	搜索引擎的利用与评价	182
6.1	搜索引擎概述	182
6.2	搜索引擎的类型及评价标准	190
6.3	国内外常用搜索引擎介绍	201
6.4	搜索引擎的功能和使用技巧	209
第7章	网络学科信息门户	217
7.1	学科信息门户的含义及由来	217
7.2	学科信息门户发展的主要形式及特点	224
7.3	学科信息门户建设的关键问题及技术	229
7.4	国外主要学科信息门户及特点	235

7.5	我国学科信息门户的建设及存在的问题	241
7.6	我国学科信息门户建设应采取的对策	248
第8章	隐性网站资源的开发与利用	252
8.1	隐性网站资源概述	253
8.2	Invisible Web 资源的利用	259
第9章	数字资源共享与文献传递服务	272
9.1	文献传递服务概述	272
9.2	数字文献传递的模式和过程	281
9.3	数字文献传递机构及其服务	292
9.4	文献传递服务供应商的选择和评价	296
9.5	我国数字文献传递服务的现状及发展对策	300
9.6	数字文献传递服务的发展趋势	316
第10章	数字信息资源的合理使用	320
10.1	合理使用制度概述	320
10.2	数字时代常见的合理使用方式	326
10.3	完善数字时代我国合理使用制度的思考	334
第11章	网络开放存取资源的利用	341
11.1	开放存取概述	341
11.2	开放存取资源的发布方式	347
11.3	网络开放存取资源的分布与利用	353
11.4	开放存取面临的主要问题	370

第1章 数字信息资源概述

随着计算机技术、网络技术、通信技术和信息存储技术的飞速发展,数字信息资源以其信息更新速度快、易检索、易利用和共享性好等特点从信息资源中脱颖而出,对社会经济、科研、教育、生活、办公、生产和管理等方方面面正产生着巨大的影响。与其他各类信息资源相比,它日益成为各种信息资源中最重要、最活跃的一种,可以说,数字信息资源是信息资源的主体,它将为人类带来新的财富、新的效率和新的生产力。

1.1 数字信息资源简介

1.1.1 数字信息资源产生的时代背景

从广义上讲,数字信息资源是文献信息的表现形式之一,是将计算机技术、通信技术及多媒体技术相互融合而形成的以数字形式发布、存取、利用的信息资源总和。从狭义上讲,数字信息资源也可称为电子文献,是指以数字代码方式将图、文、声、像等信息存储在磁、光等介质上,通过计算机或具有类似功能的设备阅读使用,用以表达思想、传播知识和积累文化的文献。它包括正式出版的电子文献(如电子图书、电子期刊、电子报纸、光盘数据库、网络数据库等),也包括非正

式出版的电子文献（如网上的各类行政报告、网上的会议资料、内部电子期刊、电子教材、信函、论坛及博客上发布的信息等）。

我们知道，在人类赖以生存和发展的自然界中，可以开发利用的物质和能源是有限的，并且这些物质和能源具有消耗性，是不可共享、不可再生的。对物质和能源的开发利用必然造成对自然的破坏和对环境的污染。在既定技术和资源条件下，其开发利用总是存在着明显的竞争关系，一部分人开发利用多了，其他人只能少开发甚至不开发。与此相比，信息是无限的、可共享、可再生的，其开发利用不仅能促进人类社会的持续发展，而且在一定程度上缓解甚至摆脱了资源短缺或枯竭的困境。随着全球信息化进程的发展，人们越来越深刻地认识到，信息也是一种资源，而且是一种十分重要的、特殊形式的资源。它与物质资源、能源资源一起被称为后工业社会（即信息社会）的三大支柱。在后工业时代，信息资源对经济社会发展的作用日益突出，已成为新的开放环境下政治、经济、文化和军事等国际竞争的焦点。近年来，信息技术的快速发展为深度开发和广泛利用信息资源创造了前所未有的条件。充分发挥信息资源对材料、能源等资源的节约和增值作用，使信息流更加有效地引导人员流、物资流和资金流，是树立和落实科学发展观、建设资源节约型社会的重要基础。目前，信息作为战略资源对经济、政治、科技、文化、军事等方面发展的重要性已得到世界各国的普遍认可，人们对资源的争夺已由占有土地、矿藏等自然资源逐渐转移到争夺信息资源。在我们大踏步地迈向信息社会的今天，信息资源重要性的观念正在深入人心。而信息社会的本义就是利用现代信息基础设施快速、方便、准确地传输信息资源。这种能够利用现代信息基础设施来

传输、记载的信息资源主要是数字信息资源。为了能清楚地理解数字信息资源的内涵，还是让我们沿着信息革命的轨迹来看一看。

自从有了人类，人们对信息的不自觉认识、开发及利用的历史就开始了。古人的记事方式是从“结绳记事”和“刻木记事”开始的，这实际上是人们早期存取信息的一种方式。人们在共同劳动和相互交往的过程中逐步创造了语言，语言成为信息交流的工具，信息从此得以口耳相传。可以说，语言能力的获得是人类历史上的第一次信息革命。

文字的出现使人类第一次突破时空的限制，使口耳相传的信息固定下来，方便了人类获取、保存和流传信息，促成了信息的完整和大量积累，并使其能够世代相传。印刷术的发明则扩大了文字的应用。公元11世纪，中国的毕昇发明了活字印刷，取代了效率较低的雕版印刷。活字印刷术传到欧洲后得到了迅速发展。1438年，约翰·古登堡（J. G. Gutenberg）发明了活字印刷机。之后，印刷术不断发展，经历了木印、石印、铅印、凹印、平印和胶印等不同的历史时期，使文字信息的大量复制成为可能。文字和印刷术的发明使文字成为信息的载体并使信息得以大量传播，这是人类历史上的第二次信息革命。

1844年5月24日，电报的发明家莫尔斯发出了“上帝创造了何等的奇迹”的电文，从此，电报通信的时代开始了。这是人类首次用电作为载体来传递信息，较之“烽火戏诸侯”时的烽火台，用“急脚递”传递信件的古代邮驿，信息传递速度大大加快。1866年，英国长达4 000公里的大西洋海底电报电缆铺设成功，这是人类有史以来实现万里瞬间通信的开端。随后，电话、无线电、广播、电视、传真机和通信卫星等通信工具相继产生，这是人类历史上的第三次信息革命。

计算机的发明则引发了人类历史上的第四次信息革命。1946年,美国莫尔电子工程学校和宾夕法尼亚大学的电子计算机设计组成功研制了世界上第一台大型电子计算机“埃尼阿克”(ENIAC),其设计思想来自数学家冯·诺依曼。在他的方案中,采用二进制的“0”和“1”来表示数据和指令,确立了数字电子计算机的体系结构。虽然计算机的发展十分迅速,半个世纪以来,其主要部件已经历了电子管——晶体管——中小规模集成电路——大规模、超大规模集成电路四个发展阶段,但冯·诺依曼的设计思想仍沿用至今。计算机的处理能力呈指数级增长,其应用也由最初的科学计算发展到信息处理,正是因为计算机的这种飞速发展和广泛应用,才使得信息资源的形式发生了变化,主要是信息资源在处理与存储时的形式发生了巨大的变化。信息资源以数字化的形式存在,也就是说,计算机实现了信息数字化的突破,所有的信息都可以用“1”和“0”这两个数字来表示,从而使信息的载体和传输介质产生了质的飞跃。计算机在加工处理信息的数量、质量、效率和速度等方面都大大超过了人脑,为了利用计算机的优点来处理信息,人们逐渐将各种信息资源数字化。之后,人们又将迅速发展的通信技术、网络技术、存储技术、多媒体技术与计算机技术结合起来,现代信息技术开始兴起,网络化、数字化的信息环境加速形成,信息资源数字化成为必然。

由信息革命的发展历程我们可以看出,每次信息革命的发生都源于记载、传输信息载体的巨大变化。数字信息资源就是在这种变化之中产生的,它是现代信息技术发展的必然产物。现代信息技术将文字、图形、图像和声音等高密度信息数字化,通过信息网络,以高速度、大容量传递到每一个办公室、教室、图书馆及家庭,使全世界的每个角落都能通过电子信息

资源实现瞬间信息共享。

1.1.2 数字信息资源的起源和发展

(1) 起源——机编文献目录数据库信息阶段

1946年2月15日,世界上第一台大型计算机“埃尼阿克”(ENIAC)诞生了,拉开了信息资源数字化的序幕。1954年美国海军兵器中心研制出基于IBM701计算机的信息检索系统,计算机首次应用于批式检索。1961年美国化学文摘社用计算机来编制《化学题录》,首次开始了电子出版物的实验。1964年美国国立医学图书馆创建的医学文献分析与检索系统MEDLARS(Medical Literature Analysis and Retrieved System)数据库问世,用户不仅可以进行多种逻辑运算,还可以从多种途径检索文献。随后,美国其他一些机构也相继使用电子计算机编辑文摘期刊,从而在客观上编制成了机器可读的书目文档,称之为文献目录型数据库。由于这种文献目录型数据库使用计算机来编辑、检索文献信息,因此我们可以将机编文献目录型数据库的研制视为数字信息资源形成的开端。

(2) 国际联机检索信息阶段

随着科技发展的需要,这种目录型数据库所涉及的专业面越来越广。除自然科学、社会科学以外,还包括经济及人文科学等。另外,还出现了与目录型数据库类似的、存取数值型信息的数值型数据库和记载事实全文的事实型数据库等。用户若要检索课题,首先必须将课题及其检索要求递交给有关的检索服务部门,检索服务部门把许多用户的课题及其检索要求加以汇总,进行批量检索,然后把检索结果邮寄给检索用户。显然,其缺点是用户不能进行人机对话,也不能及时修改检索策略,更不能及时获得检索结果。通信网络的出现和发展则弥补

了这一大缺陷,它促进了信息资源的远距离传输和交换。首先是银行业、建筑业以及商界为了及时得到准确的行情和金融方面的信息,率先使用了计算机处理技术,并进一步推动了公用通信网络与计算机系统相结合的国际联机检索系统的发展。国际联机检索就是用户使用终端设备,远距离地从国际联机检索中心迅速而准确地获取电子文献信息,使知识或信息得到广泛而有效的传播与利用,其实质就是数据库和通信的结合。从20世纪60至70年代起,许多国家还先后建立了专门从事计算机检索的机构,如美国的洛克希德公司(Lockheed)和系统发展公司(System Development Corporation);英国的目录检索服务处(Bibliographic Retrieval Service)和图书馆自动化情报服务处(British Library Automated Information Service);德国医学文献情报研究所(Deutsches Institut für Medizinische Dokumentation und Information)和欧洲太空机构(European Space Agency,简称ESA)等,这些机构都建有大量的数据库联机检索系统,向全世界联机用户提供电子信息服务。其中,比较著名的国际联机检索系统有Dialog、ORBIT、ESA-IRS、JOIS、BRS、QUESTEL、OCLC、STN等。美国Dialog系统是世界上最早和最大的专业情报检索系统,也是我国科技界广泛使用的系统。该公司自1972年起,开始为用户提供服务工作。最初由美国洛克希德导弹航空公司所属的一个情报科学实验室负责建立。1981年6月开始独立经营。1985年Dialog以3.5亿美元出售给Knight-Ridder新闻公司,成为新闻单位,现在称KR-Dialog系统。

Dialog现有全文、题录、事实及数据型数据库900多个,14亿多条记录,15TB的数据容量,近20万种数据来源,包括4.7亿条科技信息、4000多万条专利信息、超过6000万

家公司名录、20万份专业市场研究报告、2000个实时新闻资源以及由1万个新闻出版商提供的超过1亿条的新闻。专业内容覆盖自然科学、社会科学、工程技术、人文科学、时事报道及商业经济等各个领域,文献数据库有题录、文摘及全文等多种形态,科技文献包括期刊、会议录、图书、专利、科技报告、学位论文、标准、产品手册等各种类型。数据库中的数据最早可回溯至20世纪60年代,多数按周更新。

Dialog系统现拥有130多个国家约2000万个终端用户,是目前世界上最强大的国际联机检索系统^①。

(3) 光盘信息阶段

尽管利用国际联机检索系统检索到的数字文献信息具有较高的实用价值,但国际联机检索费用昂贵,一般用户难以承担,人们开始努力寻求一种低廉的存储、检索数字信息的方式,光盘存储技术则适应了这一需求。光盘是20世纪70年代出现的一种新型信息存储载体。由于光盘在存储数字信息资源方面具有存储密度高、读取速度快、容量大、体积小、寿命长、可实现随机存取和检索费用低廉等优点,因此,光盘被广泛用于存储、检索数字信息资源,并产生了一批生产系列光盘的公司,如美国UMI公司和银盘公司等。光盘记载的数字信息资源并不局限于文献信息,还包括各种软件,但可用于检索的仍以文献信息为主。

(4) 网络信息阶段

国际联机检索和光盘检索为我们提供了大量的数字信息资源,但它们各自又都有着或多或少的缺点,例如国际联机检索

^① Company Background—Dialog, A Thomson Business. <http://www.dialog.com/about/> (访问时间:2005-04-20)

的费用昂贵,检索指令复杂,只有受过专业培训的检索人员才会使用;另外,光盘检索得到的信息时效性相对较差。因此,人们渴望一种新型检索方式的出现。1993年,美国政府提出国家基础设施建设(NII)计划,兴建以Internet为雏形的“信息高速公路”。Internet的迅猛发展使网络信息资源如潮水般涌来,它已成为人类学习、工作、科学研究、商业活动和共享信息资源的重要手段,并为世界各国所接受。Internet是通过互联而形成的全球性网络,它已延伸到地球上几乎所有国家。Internet网上的所有主机都采用TCP/IP协议连接和通信,网上各种计算机都遵循该协议所规定的方式进行数据交换,其结果使得Internet存储了极为丰富的数字信息资源,堪称世界上最大的图书馆或信息资源库。Internet信息资源主要包括电子图书、电子报刊、电子新闻、电子论坛、会议资料、各种软件资料、图像文件、声音文件和电子游戏等。由于Internet起源于美国军方,并在学术界得到了迅速发展,美国政府扶持的力度较大,因此网上大部分资源是免费的,如Archie数据库、WAIS数据库、大多数大学图书馆的书刊目录数据库以及Web信息系统中的共享性数据库等。由此可见,使用Internet提供数字信息资源费用较为低廉,同时,Internet的信息更新速度快,用户得到的信息时效性比较强。此外,Internet中的WWW网络信息系统采用链式节点链接,用户检索起来不但十分方便,而且也大大拓宽了检索范围,可检索的信息除了文本信息外,还提供声音、图像等多媒体信息资源。并且,Internet还提供了多种信息服务渠道,主要有E-mail、Telnet、WWW和BBS等。为了便于充分利用丰富的信息资源,Internet上还提供各种各样的检索工具,如Archie数据库检索系统、Gopher菜单式数据库检索系统和搜索引擎等。总之,In-