

数字信息资源 检索与利用

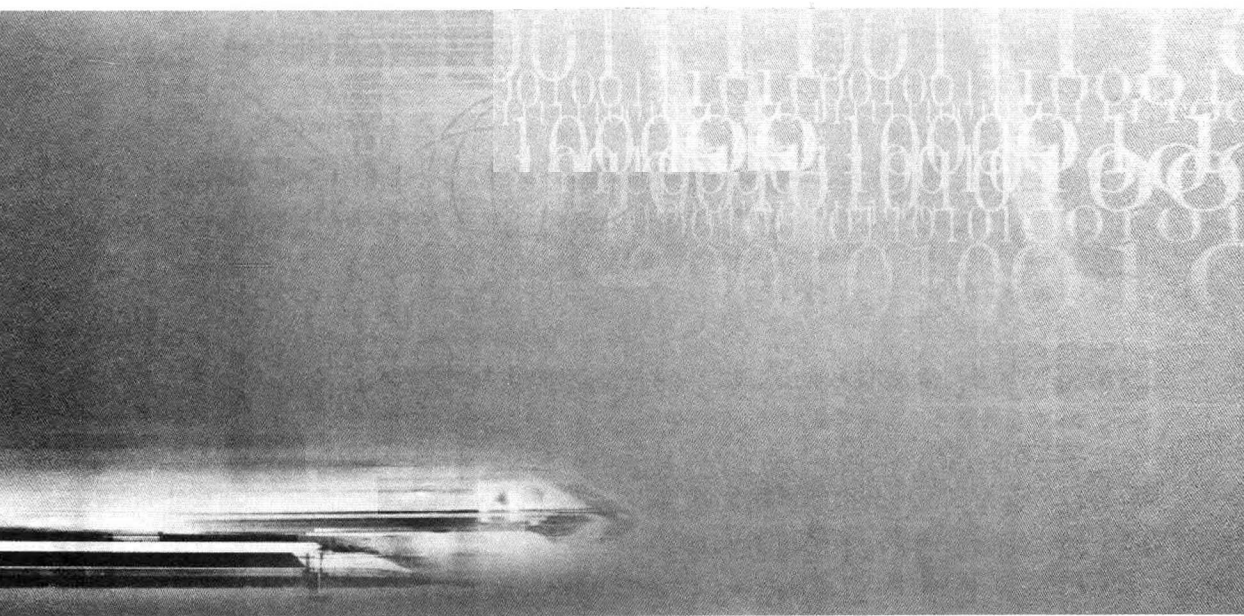
夏 红◎编著

中国科学技术大学出版社

数字信息资源检索与利用

The Retrieval and Applications of Digital Information Resources

——夏 红◎编著——



中国科学技术大学出版社

内 容 简 介

本书是作者在图书馆长期从事数字信息资源工作实践经验的基础上,依据近年来国内外数字资源检索与利用方面的新理论和新技术,并参阅有关书刊资料以及网络信息资源撰写而成的。全书分为三部分,共计七章。第一部分(第一章)系统地阐述了数字信息资源检索的基本原理、检索方法、检索技术,为基础篇;第二部分(第二、三、四、五、六章)介绍常用数字信息资源检索的基本方法、检索策略,包括检索结果处理及检索实例,为应用篇;第三部分(第七章)系统阐述了数字信息资源的综合利用以及科技论文的撰写方法,为提高篇。

本书结构清晰、内容新颖,力求给出最新、最简便、最实用的检索方法,以帮助读者提高信息检索技能以及综合运用信息的能力。本书既可作为广大信息用户查找、利用数字信息资源的实用工具书以及从事相关工作人员的研究资料,又可作为高等院校信息检索课程教学的参考用书。

图书在版编目(CIP)数据

数字信息资源检索与利用/夏红编著. —合肥:中国科学技术大学出版社,2013.10
ISBN 978-7-312-08332-2

I. 数… II. 夏… III. 计算机应用—情报检索 IV. G252.7

中国版本图书馆 CIP 数据核字 (2013) 第 239797 号

出版 中国科学技术大学出版社

安徽省合肥市金寨路 96 号,230026

<http://press.ustc.edu.cn>

印刷 安徽江淮印务有限责任公司

发行 中国科学技术大学出版社

经销 全国新华书店

开本 710 mm×1000 mm 1/16

印张 10.5

字数 225 千

版次 2013 年 10 月第 1 版

印次 2013 年 10 月第 1 次印刷

定价 25.00 元

前 言

数字信息资源(Digital Information Resources)以其信息形式的多样性、信息内容的丰富性、信息获取的便捷性、信息利用的共享性、信息传播的时效性、信息传递的交互性等特性受到人们青睐,并逐步成为人们重要的信息来源。然而,数字信息资源的不断增长以及优劣混杂使人们查找和利用信息时,常常遇到很多困扰,尤其是网络信息资源,网络的开放性、松散性,使得网络信息发布具有很大的自由度和随意性,加之缺少统一的质量控制和管理机制,信息内容十分庞杂,其信息质量表现出的不均衡性和难以控制性等特点增大了资源管理和利用的难度。数字信息产品是科学与技术的结晶,数字信息资源的收集与利用必须具备一定的知识与技能,只有掌握使用这一产品技能的人才能真正获得它的“恩惠”。

信息只有被利用才能体现其价值。如何高效、全面获取所需的优质信息资源,已成为知识学习者与研究者及广大信息用户迫切需要解决的问题。本书正是基于这种现实需求而编写的,在阐述数字信息资源检索基本理论的基础上,系统地总结了主要类型数字信息资源检索的基本方法和策略。为了便于读者学习和理解,各章的文字表述力求简明、清晰,并配以大量的图示和文字说明,以便使用者可以尽快地掌握各类数字信息资源的检索方法、途径与步骤,并能对信息资源进行有效的分析、筛选和整合。

面对数字信息资源领域的飞速发展,本书立足于对数字信息资源进行全面系统的梳理,提炼出既具有普适性又具有针对性的原理与方法,为读者快速、准确而又全面地查找数字信息资源提供有效帮助。

本书在编写过程中,参阅了许多文献及网站资料,但限于篇幅,有些参考信息未能列入参考文献之中,借此机会向所有参考文献的作者及相关网站致以诚挚的谢意!

本书在编写过程中得到了中国科学技术大学出版社的指导和大力支持,在此特表示衷心的感谢。

本书是安徽省教育厅人文社会科学研究重点项目“当代大学生网络信息利用行为研究”以及安徽省高等学校省级精品课程“化学化工文献检索与应用”资助研究成果之一。

由于编者水平所限以及编写时间仓促,加之数字信息资源领域发展迅速,书中难免有疏漏和不妥之处,敬请学界同行和广大读者批评指正,以便今后修订完善。

编 者

2013 年 7 月

目 录

前言	(i)
第一章 数字信息资源检索概述	(1)
第一节 数字信息资源的概念与类型	(1)
第二节 数字信息资源检索技术	(10)
第三节 数字信息资源检索语言	(17)
第二章 参考型数据库信息检索	(21)
第一节 索引类数据库信息检索	(22)
第二节 文摘类数据库信息检索	(41)
第三章 期刊全文数据库检索	(54)
第一节 中文科技期刊数据库	(54)
第二节 中国学术期刊网络出版总库	(57)
第三节 万方数字化期刊	(64)
第四节 Elsevier Science(SDOS)期刊全文数据库	(66)
第五节 EBSCOhost 期刊全文数据库	(70)
第四章 数字图书馆的利用	(75)
第一节 数字图书馆概述	(75)
第二节 书目信息检索	(80)
第三节 数字化图书的检索与利用	(84)
第四节 文献传递	(96)
第五章 多媒体信息检索	(103)
第一节 多媒体信息基本概念	(103)
第二节 多媒体信息检索技术	(106)
第三节 视频信息检索举要	(109)
第六章 网络信息资源的检索与利用	(118)
第一节 网络信息资源概述	(118)
第二节 网络信息资源检索	(123)

第七章 数字信息资源的综合利用	(140)
第一节 科技论文写作	(140)
第二节 数字信息资源的检索策略	(148)
第三节 检索效果评价与优化	(151)
第四节 数字信息资源使用注意事项	(154)
参考网络信息资源	(156)
参考文献	(158)

第一章 数字信息资源检索概述

第一节 数字信息资源的概念与类型

一、概述

1. 数字信息资源的概念

数字信息资源(Digital Information Resources)是以数字化形式把文字、声音、图像等形式的信息存储在磁、光、电等介质上,并通过计算机、手持终端等设备在本地或远程发布、存取、利用的信息资源。它是在计算机技术、通信技术和高密度存储技术等高新技术迅速发展并在各个领域广泛应用的环境下产生的一种信息资源形式。

2. 数字信息资源的特点

与传统型介质信息资源相比,数字信息资源具有以下特点:

(1) 存储载体的特殊性和大容量化

印刷型文献作为一个整体,其物质形态决定了文献组成部分之间的关系。而在数字化环境中,一系列相关信息资源常常汇集在同一主页下或是一个大型数据库中。数字信息资源载体具有存储密度高、存储容量大等特点,如 CD 光盘可存储 700 MB 的数据,DVD 光盘的最大容量可达 17 GB,单个硬盘的容量也有数百 GB,而磁盘阵列的容量更大。目前硬盘的单碟容量已经达到了 1 TB,可以存储 5120 亿个汉字,如果以每册图书 50 万字计算,可以存放 100 万册图书,对藏书 100 万册的大型图书馆,只需一个硬盘就可以存放图书馆的所有藏书。

(2) 资源类型多样化和信息传播网络化

数字信息资源包括不同学科、不同领域、不同地域、不同语言、动态的和静态的、正式的和非正式的、实藏的和虚拟的各种信息模式和类型,集文本、图片、动态图像、声音、超链接等多种形式为一体,包括精美的画面、优美的音乐、逼真的动画和视频图像,具体、生动、全方位地向读者展示信息的内容及特征,并可借助于网络

进行跨时空、跨行业、远距离信息传送。同时,数字信息资源的获取和交流不受时空的限制(除部分涉及知识产权保护以及内容限制访问的资源只对授权用户开放外),可随时随地存取而不受时间、地域等条件限制。

(3) 信息获取的便捷性和资源更新的时效性

数字信息资源具有很高的信息存取速度,人们可随时对其内容进行增删、修改等操作,并可根据用户的需求编制各类检索系统,使用户可以更方便、更细致地检索信息内容而不必像传统文献那样逐页查阅。用户利用检索系统可经多条检索途径获取、阅读相关数据库中的数据、图形、图像等信息,实现多层次、多角度的立体化查询。

此外,数字信息资源出版周期短、更新速度快,时刻处于一个动态变化的状态中,其时效性远远超过其他形式的信息资源。

(4) 资源利用的交互性和共享性

传统载体信息与用户之间是一种单向的关系,用户处于被动接受的地位,由于受到时间、空间、技术等各方面因素的制约,传统技术条件下的交互性不能得到充分发挥。而在网络环境下,数字信息资源系统可实现信息交流的双向传递,它不仅可以同步交互,还可以异步交互,人们在信息交流系统中可对各种多媒体信息进行发送、传播和接收等实时交互操作,受众不仅是被动的信息接受者,还是积极的大众传播的参与者,用户既是信息资源的利用者,也是信息资源开发的主体。

数字信息资源具有通用性、开放性和标准化的数据结构,为网络环境下不同形式的信息资源管理和开发利用提供了有力的支持。数字信息资源以磁性材料或光学材料为存储介质,存储信息密度高,容量大,且可以无损耗地被重复利用。不同国家、不同地区、各种服务器、各种网页及各种不同的信息资源通过节点链接起来,实现了资源的高度共享。

(5) 数字信息资源运行环境复杂、信息质量良莠不齐

与传统信息资源的保存介质相比,记录数字信息的载体一般是以光、电、磁为介质,这些介质存在易碎、易损、硬件损坏、软件故障等“载体脆弱”问题;由于信息存储格式的多样性和变化性,造成旧的数字信息不能被新的硬件设备和软件系统读出而产生“技术过时”问题。

利用数字信息资源,需要一定的软硬件设备及网络环境条件等,由于运行环境各不相同,数据库的检索界面存在着差异性和复杂性。同时,数字文献的文件格式多,如 ASCII, Tex, HTML, Word, Gif, JPEG, Word Perfect 等,不同格式对软件环境要求也不同,这无形中增加了用户检索的难度,给使用者带来了诸多不便,也在一定程度上影响了数字信息资源的使用效率。

此外,网络的开放性、松散性,使得网络信息发布具有很大的自由度和随意性,加之缺少统一的质量控制和管理机制,信息内容十分庞杂,正式出版信息与非正式出版信息交织在一起,商业信息、学术信息以及个人信息混为一体,动态新闻与稳

定信息源同处一个界面,有用信息和无用的垃圾信息混杂,甚至还有不少有害的信息掺杂其中,可谓是良莠不齐。从整体上看,网络信息资源尚处于无序状态,信息质量表现出不均衡性和难以控制性,增大了资源管理和利用的难度。

二、数字信息资源的类型

同印刷型文献相比,数字信息资源类型更为丰富。根据不同的划分标准,数字信息资源可分为不同的类型。

1. 按存储介质不同划分

按存储介质的不同可分为光介质、磁介质和电介质三种类型。光介质(Optical Medium)包括 CD、DVD、LD 等类型;磁介质(Magnetic Medium)包括软盘、硬盘、磁盘阵列、磁带等类型;电介质(Electric Medium)包括 U 盘、固态硬盘等。

2. 按数据传播的范围划分

按数据传播的范围可分为单机、局域网和广域网三种类型。单机数字资源是指存储在光盘或单个计算机上,不提供网络功能的数字资源,例如《全国报刊索引》单机版光盘、CA on CD(CA 光盘版)、《电子档案人力资源管理系统(单机版)》等,区别于网络数字资源,它不需要专门的服务器便可以正常使用该数字资源。

局域网(Local Area Network, LAN)使用方式是指用户能在机构内部浏览检索数字资源,但在机构局域网以外的网络环境中不能访问。如“广州人物数据库”(广州图书馆 <http://www.gzlib.gov.cn/>)、Gale 传记信息资源中心(上海图书馆电子资源 <http://eservice.digilib.sh.cn/>)。

广域网方式是指用户可以在任何一个拥有 Internet 的地方通过一定的身份认证或者不需认证就可以访问该数字资源,可实现不同地区、城市和国家之间的资源共享。如“黑龙江省情信息网人物志”(黑龙江省情信息网 <http://www.zglz.gov.cn/rwz.htm>)、“馆藏中文图书数字化资源库”(中国国家图书馆)等。

3. 按储存的物理地点不同划分

按储存的物理地点不同,数字信息资源又可分为现实资源和虚拟资源。现实资源是指置放于本地、本部门的数字化文献,有光盘、磁盘、磁带等多种载体形态,包括出版商出版的电子出版物、本单位制作的数字化文献以及从网络上下载到本地的信息资源。虚拟资源又称网络资源或联机资源,是指通过本单位的计算机系统及通信网络才能获取的数字化信息。

4. 按数字资源加工情况划分

按数字资源加工情况划分,可分为源数据库、参考型数据库、其他类型数字资源。源数据库(Source Database)是指用户可通过此类数据库直接获取原始资料或具体数据而不必再查阅其他信息源的自足性数据库,包括数值数据库、文本—数值数据库、全文数据库、术语数据库、图像数据库、音视频数据库等。如 CNKI 系列源

数据库——中国期刊全文数据库、中国优秀硕士学位论文全文数据库、中国重要报纸全文数据库等,中国病毒资源基础数据库(中国科学院武汉病毒研究所)。

参考型数据库(Reference Database)是指用户通过从该数据库获取的信息线索到另一信息源以获得原文或其他信息的一类数据库,包括书目数据库、文摘数据库、索引数据库等二次文献数据库。如美国《化学文摘》(CA)、《科学引文索引》(Science Citation Index, SCI)等。

其他类型数字资源包括软件资源、网页资源等。

5. 从资源提供者来划分

从资源提供者来划分,可分为商业化的数字信息资源和非商业化的数字信息资源。商业化的数字信息资源主要包括数据库商或其他机构以商业化方式提供的各种电子资源,如 Elsevier 公司的 SDOS, EBSCO 公司的 Academic Search Complete、Business Source Complete, 中国知网(CNKI), 万方数据等数字资源。这些数字资源内容丰富、数据量大,是目前图书馆馆藏资源建设中的重要内容。非商业化数字信息资源主要指机构自建的、免费开放的特色资源库以及开放获取资源(Open Access, OA)和其他免费的网络信息资源。

6. 高校图书馆数字信息资源的主要类型

在网络环境下,数字信息资源已经成为高校图书馆信息资源建设的重点,是高校图书馆在未来发展过程中面向读者服务的重要资源形式之一。

从资源建设和利用的角度划分,高校图书馆数字信息资源主要可分为电子书、电子特种文献(包括学位论文、专利、标准、会议文献等)、参考型数据库、多媒体资源库、学科导航、学习资源库、其他类型数字信息资源等。

(1) 电子书

电子书(Electronic Book, 简称 E-Book),它是一种以数字代码方式将图、文、声、像等信息存储在磁、光、电介质上,通过计算机或类似设备阅读和使用的一种电子出版物,包括电子图书、电子期刊、电子报纸等。读者可通过计算机或者阅读器阅读或下载所购买的电子图书或网站提供的电子书内容,进行在线阅读或离线阅读。

电子图书(Electronic Books)是高校图书馆重点引进及建设的数字资源之一,其内容与印刷型的图书内容基本一致。但是电子图书较印刷型纸质图书的优势在于它可提供检索、阅读及服务功能,读者可以按照自己的需求和爱好进行检索、查询、下载、标注、打印等。国内比较有影响力的电子图书数据库有超星中文电子图书、书生之家电子图书、方正电子图书等。国外电子图书的兴起和发展早于国内,而且在电子图书数据库的数量和种类上都远远超过国内。

电子期刊(Electronic Journals)就广义而言,指任何以电子形式存在的期刊,涵盖通过网络可检索到的期刊和以 CD-ROM 形式发行的期刊。电子期刊有两种类型,一种是印刷型期刊的电子版,另一种是网络版电子期刊(多为远程网络版)。国

内比较有影响力的电子期刊库有清华同方的中国知网、重庆维普、万方数据电子期刊等。国外的电子期刊库中较大型的、有影响力的学术期刊有 Elsevier Science 电子期刊全文数据库、ProQuest 电子期刊库、EBSCOhost(ASP+BSP 集成全文数据库)、Wiley-Blackwell(电子期刊数据库)、CUP(剑桥期刊在线)、WSciNet(世界科技期刊网)等,这些数据库已经成为各个高校图书馆数字资源的重点引进对象。

电子报纸(Electronic Newspaper)最初是指传统纸质印刷报纸的电子版,后来电子报纸逐渐演变成信息量更大,以及服务更加充分的网络新闻媒体,其结构与印刷型报纸基本相同,也是由刊名、刊号、出版单位或出版人、出版日期、版面数、导读栏等要素组成。相对于传统的报纸,电子报纸数据库便于管理、存储和检索。目前,国内外的电子报纸数据库有人民数据(含人民日报图文数据库)、中国重要报纸全文数据库(CCND)、人大复印报刊资料全文数据库、EBSCO-Newspaper Source、ProQuest Newspapers、Lexis 等。

(2) 电子特种文献

特种文献是指出版形式比较特殊的科技文献资料,包括学位论文、科技报告、专利文献、标准文献、会议文献、政府出版物、产品资料、其他资料(如档案资料、地图、乐谱等零散文献)等。电子特种文献资源系统包括一些专门的机构网站和数据库,如中国知网的中国重要会议论文全文数据库、国际会议论文全文数据库、中国行业标准全文数据库、万方会议文献数据库、国家科技图书文献中心(NTSL)网络版全文数据库资源、PQDD(ProQuest Digital Dissertations)博硕士论文数据库、CALIS 学位论文中心服务系统、国家知识产权局网站、美国国家标准协会(ANSI)网站、ISI Proceedings 等。

(3) 参考型数据库

参考型数据库包括书目数据库,索引、文摘类数据库,指南数据库,事实/数值型数据库等,其功能是报道文献信息的存在,揭示文献信息内容,指引用户获取原始信息的路径。

① 书目数据库

书目数据库主要存储某个领域原始文献的书目,也称二次文献数据库。书目数据库的数据结构比较简单,记录格式较为固定。在联机检索和光盘检索中,有许多书目数据库可以满足用户回溯检索和定题检索的需要。如 OPAC(Online Public Access Catalogue System)公共联机书目查询系统、CALIS 联合目录数据库、全国期刊联合目录、WorldCat(世界上最大的书目数据库)、WebPAC(新一代联机公共目录)等。

② 索引与文摘类数据库

索引与文摘类数据库是将文献中具有检索意义的事项(可以是人名、地名、概念或其他事项)按照一定方式有序地编排起来,以供检索的数据库。如《科学引文索引》(Science Citation Index, SCI)、《工程索引》(The Engineering Index, EI)、美

国《化学文摘》(Chemical Abstracts, CA)、经济学文摘型资料库(EconLit)、世界各国标准数据库(Worldwide Standards Service Plus, WWP)。

③ 指南数据库

指南数据库是对有关机构、人物等相关信息的简要描述,包括各种机构名录数据库、人物传记数据库、产品信息数据库、软件数据库、基金数据库等,亦称指示数据库。如台湾人物志资料库、《2011 全国教育机构名录数据库》、中国机构数据库(China Institution Database,简称 CIDB)、国家社科基金项目数据库、人物传记资源中心数据库(Gale-Biography Resource Center)、Peterson's Guide(彼得森研究生指南)、PIVOT 全球学者基金库等。

④ 事实或数值型数据库

事实或数值型数据库(Factual Database/Numeric Database)是将从文献中分析、概括、提取出来,或从调研、观测及统计工作中直接获得的数值型数据作为存储和检索对象的数据库,如中国公司及产品数据库、中国经济网统计数据库、CNKI 中国经济社会发展统计数据库、全国科技成果交易信息数据库、中国资讯行(China InfoBank)、Springer Protocols 数据库、Internet Address Finder(IAF)等,其数据内容准确、权威、及时,数据表现形式直观,数据检索与处理功能多样。

(4) 多媒体资源库

多媒体资源库(Multimedia Resource Database)是数据库技术与多媒体技术结合的产物,是将文本、图形、图像、声频、视频等多种媒体数据结构结合为一体,并统一进行存取、管理和应用的资源系统。如国家数字文化网多媒体资源库、爱迪科森网上报告厅、库客数字音乐图书馆、超星学术视频、知识视界、百科视频、国际数字视频图书馆、ArtBase 中国艺术品图片库等。

(5) 学科导航

学科导航(Subject Navigation)是以学科为主要分类体系,对网络学术信息资源进行分类、组织和序化整理,方便用户按学科查找相关学科信息和学术资源的系统工具。如 CALIS 重点学科网络资源导航门户、方略学科导航系统。

随着网络信息资源的不断丰富,网络中可以用于学术研究和学习的资源内容越来越多,因而,学科导航的作用显得越来越重要。

(6) 学习资源库

基于网络环境的学习资源库的建立是教育信息化的重要标志之一,为学生自主学习、教师改进教学方法提供了一个重要的信息环境。学习资源库涵盖各学科分类和各类帮助用户学习使用的资料,如图书书籍、学习课件、视频教程、课后答案等;并为学习者建立了一个交流、互动的平台,如中国高等学校教学资源网、央视教育视频资源库、新东方多媒体学习库、起点考试网、银符考试模拟题库、外研社外语资源库、世界商业领袖资料库(50 LESSONS)、大学专业课学习数据库等。

其他类型数字信息资源还包括一些专题数据库(如师范教育专题数据库、汉籍

数字图书馆)、全球产品样本数据库、中国大学生就业知识服务平台以及一些帮助读者高效利用信息资源的工具软件等。

三、数字信息资源检索的发展阶段

随着计算机技术和互联网的飞速发展,人们获取信息的来源不再仅仅是纸介质文献,数字信息资源已经越来越多,并成为重要的信息来源。信息检索也由手工检索发展到了计算机信息检索。计算机检索的成功应用,为我们更加及时、准确、全面地利用和发展人类的科研成果提供了先进的手段。

数字信息资源检索起源于 20 世纪 50 年代。1954 年,美国海军兵器中心图书馆利用 IBM 701 机开发了计算机信息检索系统,它标志着计算机信息检索阶段的开始。目前,计算机信息检索已成为用户获取信息的主要方式。纵观数字信息资源检索的发展,主要经历了脱机检索、联机检索、光盘检索和网络检索 4 个阶段。

1. 脱机检索阶段

脱机检索,又称批处理检索。1946 年世界上第一台电子计算机问世后,人们便开始研究其在信息检索领域的应用。1954 年美国海军兵器中心图书馆建立了世界上第一个计算机检索系统,是数字资源的早期检索系统,利用单台计算机,以磁带作为存储介质,进行批量或定题检索。

著名的脱机检索系统有美国国立医学图书馆的 MEDLARS(MEDical Literature Analysis and Retrieval System),1964 年由美国国立医学图书馆(National Library of Medicine, NLM)建立,实现了文献加工、检索与编制的计算机化。此外,还有美国化学文摘社发行的《化学题录》(Chemical Title)机读磁带版,人民出版社国际编辑室主任林穗芳先生曾对《化学题录》电子出版物给予了很高的评价,认为“《化学题录》电子版是第一种用计算机编辑出版的电子期刊,既是世界最早的电子出版物,也是持续出版时间最长的电子出版物之一”。

在脱机检索中,必须事先把众多用户的各种信息要求编成“用户提问档”格式,以机读形式存储在磁带上,定期地检索数据库新增加的内容,然后把命中的文献信息发给用户。其优点是实现了批处理,提高了检索速度,其不足之处是时间滞后,用户不能及时获取所需信息,检索策略不便修改,另外,其数据存取与数据通信能力都比较差。

2. 联机检索阶段

20 世纪 60~80 年代是联机检索试验和实用化阶段。1962 年,美国麻省理工学院进行了世界上最早的联机检索试验。1965 年,系统发展公司(SDC)进行了首次联机检索演示,该公司后来研制成功的联机信息检索软件 ORBIT 系统标志着联机检索阶段的正式开始。用户通过终端可用英语自然语言编排的检索策略与计算机直接进行人机对话,以查找所需的情报资料。1970 年,美国洛克希德公司创建

了 DIALOG 检索系统。美国 MEDLARS 也于 1970 年发展了联机检索系统 MEDLINE, MEDLINE 是 MEDLARS 数据库中规模最大、权威性最高的生物医学文献数据库,是目前国际医学界使用最广泛的数据库之一。

这个阶段,联机数据库被集中管理,数据库检索功能增强,能进行大批量数据处理,联机系统为全世界用户提供了丰富的电子信息资源,如著名的 DIALOG、STN 和 ORBIT 系统,其中 DIALOG 是世界上最大和历史最悠久的联机数据库信息服务系统。国际联机检索系统数据库检索到的信息资源具有实时、快速和信息追溯年代长、查准率高、实用价值较好的优点,但联机检索存在主机负担较重、对主机的依赖性过高等缺点。

3. 光盘检索阶段

光盘检索阶段始于 20 世纪 80 年代,是在计算机技术、激光存储技术等高新技术成果的基础上发展起来的。光盘检索的发展,可以说是联机检索的支持和补充。光盘是一种使用激光记录和读取信息的盘片,具有信息存取密度高、容量大、读取速度快、保存时间长、成本低等优点。光盘检索是以光盘为存储介质,利用光盘驱动器和计算机实现对光盘数据库的检索。1984 年,荷兰飞利浦公司和日本索尼公司在 COMDEX 年会上首次展出了他们的 CD-ROM 驱动器和光盘数据库,随后美国图书馆公司推出世界上第一个商品化的 CD-ROM 光盘数据库——BIBLIOFILE(美国国会图书馆的 MARC 机读目录),这是目前国际上公认的标准书目数据源。

根据使用范围,光盘检索又可分为单机版和网络版。当 CD-ROM 光盘由单机使用发展到网络,它的作用就更为明显,读者可以在图书馆、办公室、实验室甚至在家中查找图书馆的光盘数据库。光盘检索具有检索速度快、设备要求低、可随地安装、使用方便、易于操作、检索费用低廉等优点,至今仍广泛应用。

4. 网络检索阶段

网络信息检索开始于 20 世纪 90 年代初。随着卫星通信、光缆通信技术以及信息高速公路事业在全世界的迅猛发展,计算机信息检索逐步走向了全球大联网。明尼苏达大学、欧洲高能粒子协会分别推出了基于因特网的检索工具 Gopher 和 WWW。

WWW(World Wide Web, 环球信息网, 中文名字为“万维网”)将不同领域、不同学科、不同性质和不同种类的信息资源集于一体,汇集了各种电子书刊、电子新闻、电子报告、电子论坛、软件资料、图像文件、声音文件和电子游戏等,堪称世界上最大的图书馆或信息资源库。WWW 是一个基于超文本方式的信息检索服务工具,在互联网环境中,WWW 大多采用的是分布式计算机信息检索系统,通过高性能的并行检索机和大型的数据库管理系统分配、管理和调度网络中的通道和信息资源,具有模组化、并行性、共享性的特点。

网络信息资源具有内容的广泛性、信息的时效性、资源的动态性、访问的便捷

性、使用的交互性等特点,网络化检索一般都是通过搜索引擎提供多种搜索途径,其检索界面简单,操作方便,无需很高的检索技巧,不受时空限制,使得数字信息资源利用突破了传统检索方法的局限,发挥了更大的效益。

我国在数字信息资源检索方面起步较晚,但发展速度较快。20世纪80年代初,我国图书情报机构开始引进国外数据库开展计算机检索服务,先后在北京、上海、广州、武汉、成都、重庆等城市建立了近80个远程检索终端,通过国际通信网络与国外数十个大型联机检索系统进行联机检索。1986年开始,中国和美国DIALOG系统、ORBIT系统以及欧洲航天局ESA-IRS等系统实现了联机情报检索,清华大学、上海交通大学、东南大学等高校也建立了自己的国际联机终端。到90年代中期,全国有200多个联机检索终端与国外的DIALOG、ORBIT、BRS、MEDILARS、ESA-IRS、STN、DATA-STAR等国际检索系统联机。

我国各大图书情报机构在开展国际联机检索服务的同时,也纷纷开发自己的联机检索系统及数据库,向国内提供信息检索服务,有些检索系统还走出国门,面向国外用户开展数字信息检索服务。比较有代表性的有清华同方的中国知识基础设施工程(CNKI)和中国科学技术信息研究所开发的万方数据资源系统等。

四、数字信息资源检索概述

数字信息资源检索是指根据信息需求,采用一定的技术手段,通过检索系统在数据库或其他形式的数字资源中找出用户所需相关信息的过程。

数字信息资源检索是数字信息特征标识与检索需求提问标识进行匹配的过程,它包括两方面的内容:

① 信息的标引和存储

将所采集的信息按照一定规则记录在相应的信息载体上,并按照一定的特征和内容组织成系统有序的、可供检索的集合体。为了保证用户全面准确、快速地获得所需信息,需要对原始信息进行搜集、整理、著录、标引、整序,使之从分散变为集中,从无序变为有序,从不易识别变为特征化描述,以便于人们识别和查找。

② 信息的需求分析和检索

对用户所表达的信息需求进行分析和整理,并与所存储的数字信息资源进行匹配运算,简言之,就是把检索者的提问标识与存储在检索系统中的文献标识进行比较,两者一致或文献标识包含需要检索的标识时,则具有该标识特征的文献就被从检索系统中输出反馈给用户,该输出的数字信息资源即为检索初步命中的文献。

在信息资源存储过程中,需要对信息资源进行著录。信息著录即按照某种规则对某一信息资源的主要特征(外部特征和内容特征)进行分析、选择和记录。信息著录是组织检索系统的基础,是信息存储过程中的重要环节。

文献资源的外部特征是文献外部直接可见的特殊表征。如文献的题名、责任

者、序号(ISBN、ISSN等)、文种、出版事项和出处等。它是文献识别的直接依据。它的某些外部特征具有检索意义,是检索工具著录的对象。

文献资源的内容特征是指文献所含信息和知识的特殊表征,具有重要的检索意义,是检索工具描述的对象。如文献所属的学科门类、论述的主题对象、表达的基本观点等。

信息存储与检索是两个既相逆又相关的过程,信息存储是基础,信息检索是目的。

第二节 数字信息资源检索技术

由信息检索原理看,数字信息资源检索过程实际上是将检索提问词与文献记录标引词进行对比匹配的过程。为了提高检索效率,必须采用一定的检索技术。检索技术是指从数字信息资源系统中提取相关文献信息的方法和技巧,包括从概念相关性、位置相关性等方面对检索提问实行技术处理、检索词的组配形式和检索表达式的构成规则等。

常用的信息检索技术和检索方式有以下几类:

一、布尔逻辑检索

布尔逻辑检索(Boolean Operators Search)就是利用布尔逻辑算符(Boolean Operator)对检索词(短语)或代码进行逻辑组配,并指定文献的命中条件和组配次序,凡符合逻辑组配所规定条件的为命中文献。它是数字信息资源检索系统中最常用的一种检索方法。逻辑算符主要有 OR/或/+、AND/与/*、NOT/非/-,具体表述如下:

1. 逻辑“或”

逻辑“或”用符号“OR”或“+”表示,其逻辑表达式为: A OR B 或 A+B。其意义为:检索记录中凡含有检索词 A 或检索词 B,或同时含有检索词 A 和 B 的,均为命中文献,如图 1.1 中阴影所示。

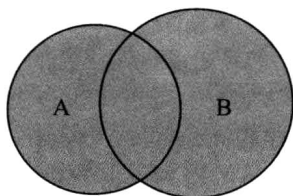


图 1.1