

21

世纪

高等医学院校教材

方平 主编

网络医学资源 检索与利用



科学出版社

www.sciencep.com

21 世纪高等医学院校教材

网络医学资源检索与利用

方 平 主 编

科 学 出 版 社

北 京

内 容 简 介

信息时代,网络技术、网络信息资源的发展一日千里。怎样快速、高效地获取网络医学资源已成为医学科研工作者亟须解决的重要课题。

本书着眼于医学信息的检索利用及教学的实际需要,详细介绍网上的各种学术资源、搜索引擎、电子图书、学术期刊、循证医学资源、WWW 文献检索系统、特种文献、多媒体资源及生物信息学资源的检索与利用;对最新的网上医学资源如可视人计划、全脑图谱、循证医学证据、生物信息学数据库等进行了详细论述。本书注重实用,对科研工作者所关心的网络学术期刊全文的获取、引文相关数据的检索与利用、如何获取医学网站学术资源等问题也详有论述。

本书适合医学专业研究生、本科生、医生和医药信息工作者阅读。

图书在版编目(CIP)数据

网络医学资源检索与利用/方平主编. —北京:科学出版社,2003.8

(21 世纪高等医学院校教材)

ISBN 7-03-011471-X

I. 网… II. 方… III. 计算机网络—医药学—情报检索—医学院校—教材 IV. G252.7

中国版本图书馆 CIP 数据核字 (2003) 第 040826 号

责任编辑:吴铁双 李 君 范 谦 / 责任校对:刘小梅

责任印制:刘士平 / 封面设计:卢秋红

版权所有,违者必究。未经本社许可,数字图书馆不得使用

科 学 出 版 社 出 版

北京东黄城根北街16号

邮政编码:100717

<http://www.sciencep.com>

保德印刷有限责任公司 印刷

科学出版社发行 各地新华书店经销

*

2003 年 8 月 第 一 版 开本:787×1092 1/16

2003 年 8 月第一次印刷 印张:19 1/4

印数:1-4 000 字数:452 000

定价:28.00 元

(如有印装质量问题,我社负责调换〈新欣〉)

《网络医学资源检索与利用》编委会

主 编 方 平

副主编 胡德华 黄碧云 刘雁书

编 者 (以姓氏笔画为序)

方 平 方国辉 刘雁书

李 珏 胡德华 莫梅琦

黄碧云 虢 毅

秘 书 李 珏

前 言

网络是传递、交流信息的一种工具,如电子信箱、网络会议、远程医疗、网上购物等,正在逐步广泛地渗透到我们的工作、学习和日常生活之中。近年来网络的迅猛发展虽然得益于商业推动,但是它在教育、科研中仍处于相当重要的地位。从网络的起源和发展来看,都与教育、科研等学术活动密切相关。网络的普及和信息的日益数字化在很大程度上改变了原有的教育和科学研究、交流方式。对于学生和学者来说,网络更是一种全新的、高效的学习工具,知识更新工具和学术交流工具。

计算机网络,特别是因特网已经成为世界上最大的信息资源宝库,其知识范围几乎涵盖人类所有学科领域。网上资源具有超越传统出版物的许多优势,如出版发行快速灵活;远程资源可与本地、本单位资源整合链接;文本资源与多媒体资源融合;检索、筛选和输出均可实现个性化,等等。网络无国界、时空的限制,使信息资源可以共享和传递,可以得天独厚地任意拓展,许多资源可以免费检索和索取,付费资源则可以在广域网或局域网内快捷方便地共享检索和利用。总之,无论你是身处繁华的城市还是偏远的乡村,网络资源宝库这个巨大的虚拟图书馆无时不以其宽厚大度的胸怀使你感受到它的无限开放和平等。

网络技术的发展面向大众用户,用户通过简便易学的操作便可驰骋网络,从中获益。然而网络资源世界浩瀚庞杂,搜索引擎的检索结果动辄成千上万;同时由于目前尚缺乏整体的过滤监督机制,网络信息流不免鱼龙混杂,泥沙俱下,欲从中获得有用的学术信息犹如沙里淘金。本书以通俗易懂的文字,介绍网络学术资源特别是医学资源及其检索语言、一般检索技术与策略、引文数据利用等内容,这些都是医学生和医学工作者在“淘金”过程中必备的基本知识。

网络资源浩如烟海。本书所介绍的网络学术资源,其选择原则主要是学术性和实用性。如网络学术期刊、循证医学资源、生物信息资源、中外引文检索数据库、PubMed、CBMdisc、MEDLINEplus 等等,都是学术界特别是医学界最关注和常用的资源,也是经过比较严格编审的、含金量高的和比较稳定的优质资源。在“重要医学网站选介”一章中,因医学网站数量巨大、内容复杂,本书篇幅所限,仅在较宽泛的医学学科领域中优选若干网站之精华,使读者通过了解这些具有代表性的优秀网站的栏目风格、内容特色和检索模式,对医学网站学术资源从不

同侧面获得较为深刻的印象。读者如需访问更多的或与自己专业有关的网站资源,可先阅读本书有关搜索引擎的章节。

网络资源的动态多变性,增加了这类著作编写的难度。本书将重点放在有关基本原理、基本知识、基本方法和重要资源的最新进展方面,这些内容既能满足广大读者所需又相对稳定。然而,这种努力恐怕也难以赶上网络资源快速发展变化的步伐,同时由于编者水平所限,本书在内容上不免挂一漏万和存在错误之处,诚请广大读者批评指正。

本书的编写和出版都得到中南大学医学技术与情报学院和中南大学医学图书馆各级领导的关心和支持,复旦大学医学院图书馆徐一新教授、中南大学医学图书馆柳晓春教授提供了许多有益的建议和帮助,在此表示最诚挚的谢意。

方 平

2003 年 3 月

目 录

第一章 网络学术资源检索概述	(1)
第一节 网络学术资源的类别、特点和评价	(1)
第二节 文献信息检索语言	(9)
第三节 一般检索技术与策略	(23)
第二章 搜索引擎	(29)
第一节 搜索引擎概述	(29)
第二节 Google 搜索引擎	(36)
第三节 医学专业搜索引擎	(46)
第三章 各类型网络文献	(60)
第一节 网络图书	(60)
第二节 网络学术期刊	(69)
第三节 专利文献及其网络资源	(75)
第四节 网络学术会议与学位论文资源	(83)
第五节 网络医学多媒体资源	(87)
第四章 常用中文网络文献检索系统	(94)
第一节 中国生物医学文献数据库	(94)
第二节 中国期刊全文数据库	(102)
第三节 全国报刊索引	(104)
第四节 国家科技图书文献中心	(107)
第五节 中国高等教育文献保障系统	(112)
第六节 万方数据资源系统	(118)
第五章 常用外文网络文献检索系统	(124)
第一节 PubMed	(124)
第二节 MEDLINEplus	(131)
第三节 EMBASE	(136)
第四节 BIOSIS PREVIEW	(141)
第五节 《化学文摘》光盘数据库	(145)
第六节 Ovid 数据库系统	(151)
第七节 DIALOG 系统	(154)
第八节 ISI Web of Knowledge	(158)
第六章 循证医学及其资源检索	(175)

第一节 循证医学概述	(175)
第二节 循证医学证据来源	(185)
第七章 生物信息学数据库及其检索	(193)
第一节 生物信息学数据库	(193)
第二节 生物信息学数据库查询与检索	(204)
第八章 引文检索	(219)
第一节 引文索引及其相关数据的利用	(219)
第二节 ISI Web of Science	(225)
第三节 国内可检索引文的数据库	(231)
第九章 重要医学网站选介	(240)
第一节 综合性医学网站	(240)
第二节 预防医学网站	(253)
第三节 基础医学网站	(261)
第四节 临床医学网站	(266)
第五节 药学与药理学网站	(272)
第六节 护理学网站	(279)
参考文献	(283)
附录一 MeSH 树状结构表主要类目	(286)
附录二 MeSH 副主题词使用范围	(290)
附录三 MeSH 副主题词扩展表	(296)

第一章 网络学术资源检索概述

第一节 网络学术资源的类别、特点和评价

网络学术资源(Network Academic Resources)是通过计算机网络可以利用的各种学术资源的总和,也可以理解为计算机网络中存在的各种学术资源的总和,即各类局域网或广域网存在的各种学术资源的总和。作为文献信息在网络环境中的表现形式之一,它包括以下几个部分:一是传统印刷文献的网络版,它是已经存在的传统文献经过电子扫描或人工输入等方式进入网络的;二是出版发行的电子形式产品的网络版,例如将内容用数字符号的形式存储于磁带、磁盘和光盘等载体,然后放到网上;三是依托于计算机网络环境,以电子形式在网络环境中自由传递与存取的各种学术资源。总之,网络学术资源不仅仅是传统出版物的电子化,还是各种学术资源直接在网络上出版、流通和传播。

网络学术资源极其丰富,其内容涉及农业、工业、生物、化学、数学、天文、地理、航天航空、计算机、医学、政治、法律、文学、历史等所有专业领域。它是学术资源重要组成部分之一,是人类宝贵的精神财富。

一、网络学术资源的类别

1. 根据网络学术资源传播范围分类

根据网络学术资源传播范围,可以分成光盘局域网学术资源、联机网络学术资源和因特网学术资源。从它们的应用范围和使用情况,我们不难发现因特网学术资源是网络学术资源的发展趋势。

(1) 光盘局域网学术资源:光盘是 20 世纪 80 年代发展起来的一种新型信息存储介质,以其存储容量大、读取速度快、信息类型多样等诸多优点而得以广泛应用。自 1985 年世界上第一个商品化光盘数据库——Bibliofile(美国国会图书馆的 MARC 机读目录)问世以来,全世界已生产各类光盘数据库及其他学术资源上万种。随着网络技术的发展,特别是光盘塔、大容量硬盘和磁盘阵列的出现和广泛应用,光盘网络检索系统在公共图书馆和大学图书馆纷纷建立起来,使得光盘的多用户检索和共享成为现实。现在越来越多的光盘局域网学术资源发展成为广域网或 Internet 学术资源,这样,大大提高了光盘局域网学术资源的利用率。

(2) 联机网络学术资源:20 世纪 60 至 70 年代,欧美等发达国家相继建立计算机联机信息检索系统,如美国的 Dialog 系统、ORBIT 系统和 MEDLARS 系统,德国的 STN 系统等,向世界范围内有限的用户提供信息检索服务。受 Internet 的巨大冲击,这种集中式的联机信

息检索系统的局限性日益突出,并严重制约了它的发展。近年来,世界知名的联机信息检索系统如 Dialog、MEDLARS、STN 等纷纷建立自己的 WWW 服务器,开发 Internet 接口,改善用户界面,增加服务项目和内容,将其服务对象从原来有限的用户扩展到世界各地。

由于这些联机检索系统在信息加工上的优势和在信息服务方面的独到之处,使得联机网络学术资源以其加工标引规范、检准率高、数据库涉及学科范围广、专业性强而逐渐成为 Internet 一种重要的学术资源。

(3) 因特网学术资源:随着因特网的迅速发展和普及,因特网学术资源成为近年来发展最为迅速的一类网络学术资源。因特网学术资源不仅内容丰富多彩,而且表现形式多样。它不仅包括目录、索引、网站、网页乃至全文,还包括声音、动画、图像和其他多媒体信息;不仅包括万维网(WWW)、FTP、Telnet、Gopher、WAIS 等网络工具上的学术资源,还包括新闻组(USENET/newsgroup)、电子邮件群(LISTSERV)、用户邮件群(Mailing List)、电子公告板(BBS)以及电子论坛等网络工具上的学术资源。

这里所指的因特网学术资源是一个狭义的概念,是针对光盘局域网学术资源和传统的联机网络学术资源而言的。其实,光盘局域网学术资源和传统的联机网络学术资源日益融入因特网学术资源。

2. 按学术交流的方式分类

(1) 非正式出版的学术资源:是指流动性和随意性较强、信息量大、信息质量难以保证和控制的动态性资源,如电子邮件、电子论坛、电子学术会议、电子公告板、新闻组等工具上的资源。

(2) 半正式出版的学术资源:又称“灰色”学术资源,是指受到一定知识产权保护但没有纳入正式出版系统和交流渠道的学术资源。如从各种学术团体和教育机构、科研单位、国际组织和政府部门、企业和商业部门、行业协会等单位的网址或主页上,可以查询在正式出版系统中所无法得到的“灰色”学术资源。

(3) 正式出版的学术资源:是指受到一定的知识产权保护、信息质量可靠、利用率较高的学术资源,如各种网络数据库、电子期刊、电子图书、电子参考工具书、电子报纸、专利信息、联机馆藏目录等。

3. 按网络学术资源加工程度分类

按网络学术资源加工程度,可以分为网络学术资源指南与搜索引擎、联机馆藏目录、网络数据库、电子出版物、参考工具书和其他动态信息。

(1) 网络学术资源指南与搜索引擎(详见第二章):网络学术资源指南是一种等级式分类目录,类目按一定的分类体系组织,排列方法多种多样。用户通过逐层浏览类目、逐步细化的方式来寻找合适的资源。人工参与网页信息的收集、筛选、标引、编排,收录的资源质量高。常见的有美国国会图书馆编辑的 WWW Meta Index and Search Tools、美国伊利诺大学的国家超级电脑应用中心编辑的 Internet Resources Meta Index,以及 Yahoo、Whole Internet Catalog 等。

搜索引擎则是自动搜索采集网页信息、自动标引,提供布尔逻辑检索、短语检索、自然语言检索等多种查询方式。除了常规搜索引擎外,还有多元搜索引擎,以及一些专门搜索 Gopher 地址、FTP 地址、USENET 以及 E-mail 地址的搜索引擎。常见的常规搜索引擎有 Google、Alta

Vista、Infoseek、Lycos 等。

值得一提的是,搜索引擎和资源指南的界线渐趋模糊,多数网络检索工具同时具有关键词检索和目录浏览功能。

(2) 联机馆藏目录:联机馆藏目录是各图书馆现实馆藏文献的检索系统,在揭示馆藏文献内容和提供检索、馆藏利用以及馆际互借、资源共享等方面发挥着非常重要的作用。

目前全球 600 多所著名的公共图书馆、大学图书馆以及 400 多个学术机构将其馆藏目录通过因特网向公众免费开放。这些图书馆馆藏目录检索系统称为“联机公共检索目录”(Online Public Access Catalog, OPAC)。例如,美国国立医学图书馆(NLM)(<http://www.nlm.nih.gov>)在网上提供馆藏目录检索。我国从 1998 年开始建设中国高等教育文献保障系统(China Academic Library and Information System, CALIS)(参见第四章第五节),通过中国教育科研网(CERNET)提供中外文书刊联合目录和七个地区级书刊联合目录检索、馆际互借和文献传递服务。世界上最大的联机图书馆文献数据中心 OCLC 系统也联入因特网,它拥有 3700 余万条编目记录(每年还以约 200 万条编目记录递增)和世界各地 25000 余所成员馆的 6 亿条馆藏记录;还收录了 800 万篇文献,每天有 5000 篇文献加入该数据库,内容涉及科技、医学、人文科学和社会科学。越来越多的图书馆在提供联机馆藏目录的同时,开始提供流通服务,读者可联机提出借阅请求,图书馆可将图书或文献寄给读者。

(3) 网络数据库:网络数据库(Web Database)主要是指出版商和数据库生产商在因特网上发行的出版物和数据库。它可以是电子图书、电子期刊、电子报纸等一次文献数据库,也可以是文摘、索引、目录等二次文献数据库。

网络数据库通常经订购后直接通过因特网或经本地镜像站点访问检索,同时依托网络发行传递的快捷方便,日益将信息检索与原文传递和最新文献报道等服务融为一体。

例如,由美国 Carl 公司出版发行的 Uncover 期刊数据库是当前世界上规模最大、内容更新最快的网络期刊数据库之一。Uncover 收录自然科学和社会科学 18000 种期刊,除了提供检索外,还提供“最新文献报道”和原始文献传递服务。“万方数据资源系统”(<http://www.wanfangdata.com.cn>)是建立在网络上的大型综合性信息资源系统(参见第四章第六节),以科技信息为主导,同时涵盖经济、文化、教育等相关信息,该系统在网上提供 80 多个数据库的检索。中国期刊网(CNKI,<http://www.cnki.net>)(参见第四章第二节)自行研制开发了四大系列数据库,一是期刊系列,包括题录数据库(免费)、题录文摘数据库和全文数据库;二是报纸系列,包括中国重要报纸专题数据库(现阶段免费使用);三是会议论文系列,有中国会议论文数据库;四是博士和硕士论文系列,有中国优秀博士/硕士论文全文数据库。

(4) 电子出版物:电子出版物主要包括网络图书、网络学术期刊和网络报纸。

网络图书是网络中一种重要的学术资源(详见第三章第一节)。如《Grays Anatomy of the Human Body》(《格氏人体解剖学》)是世界上出版历史最久、最著名的人体解剖学专著,Bartleby 公司于 2000 年在网上推出该书经典的 1918 年第 20 版的网络版(<http://www.bartleby.com/107>),向广大医学生和医学科研人员提供免费阅读服务。又如耶鲁大学医学院推出的纯网络版《Yale Heart Book》(《耶鲁心脏病学》)(<http://www.med.yale.edu/library/heartbk>)全书共分 8 个部分,29 章,涉及心脏功能、心脏病的危险因素和诊断步骤、各类心脏病、特殊人群心脏病及治疗方法等内容。读者可免费阅读该书的全文,也可下载打印

该书的 PDF 格式文件。

网络学术期刊是指以电子文件格式发行,并且能通过计算机网络传播、检索和浏览的期刊。它既包括只以电子文件格式发行的期刊,也包括传统印刷型期刊的网上电子版(详见第三章第二节)。

目前全球上网的报纸已有几千种。我国的网络报纸是近几年起步的。目前已有近百种报纸进入因特网成为网络学术资源,大多是全国性的和地方有影响的报纸,如《健康报》、《人民日报》、《光明日报》等。

(5) 网络参考工具书:参考工具书是各类知识查考性图书,如词典、百科全书、年鉴、手册、地图、名录等。具有参考性、概述性和易检性等特点。随着 Internet 的发展,网上涌现出越来越多的各种类型的网络参考工具书(详见第三章第一节)。与印刷型参考工具书相比,网络参考工具书的内容更丰富、使用更方便、数据更新颖。如《Encyclopedia Britannica Online》(<http://www.eb.com>)是著名的《不列颠百科全书》网络版。在保留原百科全书的质量和特色的基础上,又增加了许多新的功能:强大的关键词检索功能、按主题字顺浏览功能、因特网指南(由百科全书的编辑们搜集、选择和推荐一些最好的因特网站点)和一些选自最著名杂志和报纸上的文章等。该工具书两周更新一次,需付费使用,但任何读者都可以申请 14 天的免费试用。又如《Dorland's Medical Dictionary》(<http://www.merckmedicus.com>)是 WB Saunders 公司出版的《Dorland 图解医学词典》的网络版,提供 10 万多条医学专业词汇的音标、定义、相关词汇链接等,免费使用。还有著名的《默克诊疗手册》(《The Merck Manual of Diagnosis and Therapy》(<http://www.merck.com/pubs/mmanual>),是 Merck 公司出版的使用最广泛的临床医疗手册,有 100 多年的历史。网上免费提供 1999 年出版的第 17 版的全文,共 23 篇,308 章,涉及几千个主题。可以按章节浏览,也可用关键词查询。

4. 按信息服务方式分类

(1) 万维网学术资源:World Wide Web(WWW,即万维网)万维网采用客户机/服务器体系,用户能够轻松、方便地利用分布在网络上的各种信息资源,实现互联网环境下分布式 WWW 检索体系。WWW 利用超文本标记语言(Hypertext Markup Language, HTML)在文件中标记链源和链宿,其中链宿用“统一资源定位器”(Universal Resource Location, URL)来描述。如果用户对于链宿的内容感兴趣,只需用鼠标点击链源,系统就能自动获取链宿 URL 中的信息,利用指定的协议到指定的主机地址和路径中取出指定的文件,并在浏览器中显示。

浏览器是 WWW 服务的客户端软件,如 Netscape Navigator 或 Internet Explorer,它们可以阅读 HTML 文件,可以通过 HTTP(超文本传输协议)和 TCP/IP 协议向 URL 对应的服务器发送和调用特定的信息资源。

WWW 服务器运行服务器软件和存储各种信息资源,这些信息资源是采用 HTML 语言编写的。WWW 服务器在接受客户机请求后调出指定文件,通过 HTTP 和 TCP/IP 协议向客户机送回指定资源,由客户端的浏览器显示。

(2) FTP 学术资源:FTP(File Transfer Protocol)称为文件传送协议,是因特网上历史悠久和应用广泛的网络工具。它允许人们通过协议连接到因特网的一个远程主机上读取所需文件并下载到自己的计算机上,传送的文件可以是文本、图像、声音、多媒体、数据库和可执

行的二进制代码。

从某种意义上来说,FTP 就相当于在网络上两个主机间拷贝文档,它曾经是互联网信息流量的主力,目前仍然是发布、传递软件和长文件的主要方法。

FTP 仍是因特网上重要的信息源之一,目前有大量的 FTP 资源库。但要查找所需文件的主机地址、目录路径和具体文件名,就需要特定的检索工具。

(3) Telnet 学术资源:Telnet 是因特网的远程登录协议,允许用户将自己的计算机作为某一个因特网主机的远程终端与该主机相连。从而使用该主机的硬件、软件和信息资源。许多机构都建立了可供远程登录的信息系统,如各类图书馆的公共目录系统、信息服务机构的综合信息系统、政府和公共事业部门的信息系统、商业化数据库系统等。

(4) USENET/Newsgroup 学术资源:USENET 是因特网上的一种应用软件,用于提供新闻组(Newsgroup)服务。在这个服务体系中,有众多的新闻服务器,它们作为因特网主机运行的服务器(News Server)软件,接收和存储有关主题的消息,供自己的用户查阅。用户可在自己的主机上运行新闻组阅读器软件(News Reader),申请加入某个新闻组,并从服务器中读取新闻组消息或将自己的意见发送到新闻组中。用户可查阅别人的意见并予以回复,由此反复形成讨论,所以新闻组又称“电子论坛”。

众多新闻组每天都发布大量新信息。用户使用新闻组时首先要订阅感兴趣的新闻组,选择要阅读的新闻和感兴趣的消息,然后将它们下载到自己的计算机上阅读,并针对所阅读的消息进行回复,回复可送给消息的原始发布者或新闻组的所有成员,用户也可将自己的新消息张贴在新闻组上。而且一些新闻阅读器允许用户根据新闻组中的关于某个具体问题的线索进行检索。

(5) Listserv/Mailing list 学术资源:互联网上进行交流和讨论的主要工具有三种:USENET/Newsgroup(新闻组)、Listserv(电子邮件群)和 Mailing list(用户邮件群)。这三种工具的原理和使用方法非常相似,均用于网络用户间的信息交流。但差别如下:

第一,新闻组和电子邮件群往往涉及较为广泛的内容,对参与用户也没有限制,但用户邮件群通常涉及较专门、甚至有争议的议题,对参与的用户有一定的限制。

第二,通过连入新闻组服务器即可订阅新闻组,无须专门的订阅手续,而电子邮件群和用户邮件群均要求用户履行专门的订阅手续,用户只有得到答复后才能加入。

第三,新闻组的信息存放在服务器上,用户要通过专门的客户软件阅读所需信息,而电子邮件群和用户邮件群要将信息传送到用户的电子邮件地址。

第四,新闻组和电子邮件群大多是自动管理,用户邮件组多为人工管理。

电子邮件群和用户邮件群有管理地址和邮件地址。其中,管理地址负责接受订阅、暂停发送、取消订阅等信息,通常电子邮件群在网络地址前加“LISTSERV@”,用户邮件群在网络地址前加“request@”。邮件地址负责接收需要发送的具体消息。加入电子邮件群和用户邮件群必须申请订阅,即向其管理地址发送一个订阅申请,待收到同意订阅的回复后才能通过其发送和接收消息。

电子邮件群和用户邮件群都是一对多的交流工具,可以将信息广播给所有成员,也可以只传送给某个成员。

二、网络学术资源的特点

因特网是全球最大的计算机网络,也是全球最大的信息网络。它是由许多网络构成的网络群体,同时也是一个信息资源的宝库,各种各样的信息分布式地存储在网络的各个主机之中,包罗万象。因特网的出现,无论从广度和深度上,改变了传统信息处理、传递、检索与利用的方法和方式,使信息内容更加丰富,信息来源更加广阔,信息传递更加迅速,信息检索更加方便。具体地讲,网络学术资源主要有如下一些特点。

(1) 信息量大,信息来源广泛:网络已经成为继电视、广播和报纸之外的第四媒体,是学术资源存储和传播的主要媒介之一。网络学术资源数量十分巨大,每天以 10 万网页、2000 万个单词的速度扩增。目前网上拥有 600 多个大型联网图书馆、400 多个联网的学术文献库、2000 多种网上杂志、900 多种新闻媒体的网络版、50 多万个网站,总计近 100 多万个信息源正在为人类提供信息资源交流和共享。而且这些信息源十分广泛,既有政府部门、大学院所、科研机构、学术团体、行业协会,更有大量的公司、企业和个人。

(2) 信息层次众多,品种多样:网络学术资源层次众多,有零次信息、一次信息、二次信息、三次信息等;有文本、图像、表格、超文本等静态信息,还有声音、动画等多媒体信息;包括各种电子书刊、书目数据库、联机数据库、软件资源等,因此网络学术资源是多媒体、多语种、多类型信息的混合体。

(3) 信息内容广泛,质量不一:网络学术资源几乎包含了科学、政治、商业、娱乐等所有领域,在网上不仅有题录、摘要甚至全文,还有许多专门数据库及虚拟图书馆。但是,这些网络信息大都没有经过严格的审查,信息发布具有很大的随意性和自由度,缺乏必要的过滤监督和质量控制。正式出版物与非正式出版物交织在一起,科技信息、学术信息、商业信息、个人信息与一些暴力、色情等污染信息混为一体。信息质量良莠不分、参差不齐,既有大量国际水平的研究成果,又有许多虚假信息甚至错误信息,给利用有价值的网络信息带来极大的不便。

(4) 信息资源分散无序:网络学术资源的分散表现在信息没有一个中心点,也没有全面性的权限。通过一篇文献可以连接到更多相关或相似的文献;同样,这篇文献也可能是从另一篇文献连接而来的。这种前所未有的自由度使网络学术资源的共建和共享变得潜力无穷,然而也使信息资源处于无序状态,而且“海量”的信息和快捷的传播加剧了这种无序状态,许多信息资源缺乏加工和组织,只是时间序列的信息堆积,缺乏系统性和组织性。

(5) 资源的时效性强,变化迅速:在网上能得到最新的资料以及某一学科或某项科研的最新动态,能检索到最近的(甚至当日)的文献。某些科研人员往往在正式发表之前,就将其研究成果放到网上。不仅如此,一些期刊还发行免费电子版文章,这些电子版文章比正式印刷出版的期刊可提早约两个月与读者见面。同时,在科研领域广泛应用的通信讨论组以及网络新闻也是获得最新科技信息(其中有大量的零次文献和零次情报)的有效途径,我们也可以免费获取每周一次的美国最新专利公告。

网络学术资源是一个动态系统,许多信息随时间的推移不断更新,地址变化较大。有的 Web 页几周就更新一次,某些重要的站点,其信息是每月、每周甚至更频繁地更新。

(6) 交互性强:与传统的媒介相比,交互性是网络信息传播的一大特点,具体体现在它具有主动性、参与性、交互性和操作性。以往,信息流动是单向的,网络信息源则要求人们自己主动到网上数据库、电子图书馆中查找自己所需的信息,还要积极地向网络上输送信息或通过电子信箱交流信息。网络信息流动是双向的、互动的。

(7) 信息检索快捷、方便:在网络中信息传递快捷,检索迅速方便,大大提高信息工作的效率。网络信息检索方式(途径)很多,可以通过 web 检索信息,也可以通过 BBS 检索信息,还可以用 Archive、Gopher 检索;检索工具也很多,如供检索 web 信息的各种搜索引擎就有上百个。此外,网络信息检索没有时间和空间的限制。用户坐在家里或办公室的计算机前,随时都可以检索到来自世界上任何地方的信息。

三、网络学术资源的评价

1. 评价网络学术资源的必要性

人们在利用印刷型文献信息资源时,一般要对众多相关主题的信息资源进行评价和选择。在利用网上信息时同样也会面临这一问题。近年来,随着因特网的急剧膨胀,人们感到面对网上的海量信息变得无所适从,在网上获取和选择信息更加困难。

(1) 获取网络信息的需要:网络学术资源爆炸式增长,使得人们从中获取有用信息的难度越来越大。在因特网上利用搜索引擎查找资料,任意输入一个关键词,都可以获得几十、几百乃至成千上万条信息,从中找出真正有价值的信息,犹如大海捞针,以致产生了“因特网信息检索定律”:在因特网上您总能找到(甚至只能找到)您所不需要的东西。

(2) 选择网络信息的需要:网络提倡的开放和自由,使得信息的发布和传播缺乏严格的审核和过滤机制。在非网络学术资源中,图书、期刊、报纸的出版都有一整套严格的审稿和出版制度,信息内容需经编辑严格审阅,予以把关;书目、索引等工具书的出版还需相关学科的专家学者审阅筛选,通常还会将审稿意见反馈给作者,以便对稿件进行修改,有时,这个过程还要反复进行多次。而网络使用者可以集作者和出版者双重身份于一身,随意发布任何信息,缺乏编辑出版这一至关重要的质量控制环节,致使信息内容良莠不齐、真伪难辨,信息污染程度日益加深,人们选择信息更加困难。

2. 开展网络学术资源评价的目的

(1) 有助于用户选择、利用网络信息。通过对网络学术资源的评价,了解网上相关学科、专业、主题领域内的学术信息的分布及质量水平等情况,从而为有关信息的取舍提供判断依据,以便在最短时间内,以最快的速度,帮助用户选择或直接为其提供最有针对性的信息。

(2) 通过对网络学术资源的评价,掌握各领域中的优秀网站网页,日积月累,形成各领域优秀网站网页群,以最终确定相关学科、专业、主题领域内最常用的“核心网站/网页”。

3. 网络学术资源评价指标

尽管可以利用印刷型环境下的文献评价指标来评价网络学术资源,但由于网络学术资源具有高度离散、无序、动态、对站点的软硬件和服务等客观条件的依赖性较大,以及其多媒体形式等特性,故应对此进行修正,提出真正适合网络学术资源的评价指标。一般包括以下

几方面:范围、内容、用户对象、图形和多媒体设计、易用性、价格等。具体分述如下:

- 网站的目的。包括该资源设立的目的、是否包含明确声明、信息资源本身是否已经实现了预期目标、特定的用户群包含哪些、信息是在什么程度和层次上提供给用户的、信息能否满足目标用户的需求,这是评价网络资源的首要因素。

- 站点收录范围。广度、深度是评价网络学术资源的重要标准。广度即是否涉及所覆盖主题领域的各个方面,深度即所提供的信息是包含原始信息还是只提供信息线索。此外,该类指标还包括所提供信息的时间跨度、信息类型或表现形式。

- 信息内容质量。这是以内容为导向的网络学术资源评价中最重要、最基本的标准,包括学术水平、可信度、时效性、内容的连续性等方面。具体为:①准确性:包括信息是否准确可靠,评价者可以通过与其他相关资源进行比较或以自己的专业知识进行鉴别和评价;②权威性:包括该信息资源是否为有声望的机构或专家制作或拥有,作者是否为该领域的专家等;③新颖性:包括信息内容是否经常更新,更新周期为多长等;④独特性:该资源是否可以通过其他形式获取(如其他站点、CD-ROM 或者印刷本),该资源在形式上具有何种特别之处等;⑤相关链接:外部链接是否明晰,链接是否能定期更新以保证链接路径正确等。

- 图形和多媒体设计。包括图形和多媒体设计是否适合于该资源的目的、是否美观、是否能增强信息资源的表现力、是否与信息内容紧密联系等。

- 网站易用程度。信息资源是否便于使用是网络学术资源评价与印刷型资源评价的不同之处。易用程度的标准主要包括:①用户界面的友好性;②所需计算机运行环境;③资源是否按照一定的逻辑方式组织以便于定位,是否提供了有用的搜索引擎,检索结果输出是否方便(如按相关度排列)等;④资源与用户间的交互性如何,是否提高了该资源价值等。

- 稳定性和连续性。包括是否可以可靠和稳定地获取该资源等。站点的稳定性和连续性至关重要,一般说来,大型机构由于有充足的资金和人力支持,比较正规、稳定;而个人站点的稳定性、连续性较差,生命周期也较短。

4. 网络学术资源的评价方法

网络学术资源评价方法主要有第三方评价法、用户评价法以及由文献计量学引申和发展起来的网络计量学。

(1) 第三方评价法:主要是相对于网络学术资源的发布者(所有者)和网络学术资源用户而言的,目前的主要形式有:①商业性的专业网络资源评价网站,如 Magellan Internet Guide(<http://www.mckinley.com>)、Lycos Top 5% (<http://point.lycos.com/categories>)、Argus Clearinghouse(<http://www.clearinghouse.net>)等。评价的范围多侧重于综合性网络资源,面向普通网络用户,所选择的评价指标体系包括日访问量、网站设计的感官效果等,注重网络资源的形式而不注重信息内容。②由图书馆所提供的网络资源评价服务,一般针对学术信息资源评价,具有专业性,采用的评价指标体系多侧重于信息内容,且考虑网络信息的权威性、学术性,是专为科学研究而服务的,如 ADAM (<http://adam.ac.uk>)、EELS(<http://www.ub2.lu.se/eels/>)、OMNI(<http://omni.ac.uk/>)、SOSIG(<http://www.sosig.ac.uk/>)等。

第三方评价法是目前较为普遍的网络学术资源评价方法,但也存在着一些缺点:①评价

指标体系的选择、相关指标权重和赋值以及评价的方法与过程往往具有不可克服的主观性,从而影响评价的客观性;②网络信息的易变性和动态性使得网络学术资源的评价工作往往滞后于实际情况的变化;③第三方制定的网络学术资源的质量指标仍难以考虑到用户的个性化与特殊化信息需求,由此会导致评价结果的适用性问题。

(2) 用户评价法:主要是由有关网络资源评价的专业机构向用户提供相关的评价指标体系和方法,由用户根据其特定信息需求从中选择符合其需要的评价指标和方法。这有助于用户搜集完全符合自身特定需要的网络学术资源,提高网络学术资源过滤的质量。用户评价法虽然克服了第三方评价法的静止性和滞后性同网络学术资源的动态性和易变性之间的矛盾,但是由于需要用户依照评价指标和评价方法对每一个网络学术资源实体(网站、网页、网络文献等)进行鉴别和评价,反而增加了用户的负担,使得用户在因特网信息海洋中无法快速发现高质量的信息。同时,由用户而不是由专业机构承担资源收集的职责,在一定程度上影响了资源收集的全面性。用户评价所采用的方法多为定性评价方法。

(3) 网络计量学:是根据网络学术资源自身特征和规律对网络学术资源进行定量评价的一种评价方法,称之为“Cybermetrics”或“Webmetrics”,即网络计量学。网络计量学在一定程度上克服了第三方评价和用户评价的主观性和偏向性,为人们提供了一种系统、客观、规范的数量分析方法,评价结果更加直观、具体,是网络信息评价的发展方向。近年来,国外已开始探讨定量研究网络特征的方法,提出了通过计算网络资源被检索或被引用的次数来评价网络学术资源。

引文分析法是评价期刊质量的经典方法。同样,网络学术资源也可采用“Web 引文索引”来评价。Web 网站中的链接可被看做类似于印刷型出版物中的引文,可通过计算其相关的数量指标来计算网站的相对质量。与计算期刊影响因子时必须考虑期刊年度发文量相类似,也必须考虑网站规模等因素,另外还要考虑站内链接和站外链接等区别。

第二节 文献信息检索语言

一、文献信息检索语言概述

(一) 文献信息检索语言的涵义

文献信息检索语言是文献信息检索系统(包括传统的文献数据库和网络信息检索工具)中的标识系统,提供检索的出发点,如著者、分类号、主题词、关键词等,又称文献存储与检索语言、索引语言、标引语言等,是由人工设计或计算机自动生成的。

文献信息检索语言的实质是表达文献信息和网络信息内容的概念及其相互关系的概念标识系统。在表现形式上,它既可以从自然语言中或专业文献中抽取出来并予以规范化的一套词汇,也可以是代表某种分类体系的一套分类代码,还可以是代表某一类事物的某一方面特征的一套代码(如代表化合物的多种代码),等等。它们可用于对文献和网络信息的内容进行主题标引、逻辑分类或特定信息的揭示和描述,例如《汉语主题词表》、《医学主题词表》、《中国图书馆分类法》等,都可称为文献信息检索语言。