



卫生部“十一五”规划教材 全国高等医药教材建设研究会规划教材
全国高等学校医学研究生规划教材

医学信息搜集的 途径与方法

主 编 聂绍平



人民卫生出版社
PEOPLE'S MEDICAL PUBLISHING HOUSE



卫生部“十一五”规划教材 全国高等医药教材建设研究会规划教材

全国高等学校医学研究生规划教材

医学信息搜集的 途径与方法

主 编 聂绍平
副主编 朱培元 张政宝

编者(以姓氏汉语拼音为序)

蔡继峰(中南大学湘雅医学院)
代成波(广东省人民医院)
蒋晨曦(首都医科大学)
蒋柳明(温州医学院)
李 博(中国中医科学院)
刘明杰(中国疾病预防控制中心)
刘文生(哈尔滨医科大学)
聂绍平(首都医科大学)
阙 斌(首都医科大学)
石 翡(汤姆森路透科技信息集团)

田 颖(首都医科大学)
徐远飞(中国医院知识仓库编辑委员会)
于洪伟(哈尔滨医科大学)
岳卫平(汤姆森路透科技信息集团)
张 帆(汤姆森路透科技信息集团)
张 铭(首都医科大学)
张政宝(南方医科大学)
朱 伟(泸州医学院)
朱培元(南京军区总医院)

编写秘书

张 铭(兼)



人民卫生出版社
PEOPLE'S MEDICAL PUBLISHING HOUSE

图书在版编目 (CIP) 数据

医学信息搜集的途径与方法/聂绍平主编. —北京:
人民卫生出版社, 2008. 9

ISBN 978-7-117-10446-3

I. 医… II. 聂… III. 医药学-情报检索
IV. G252.7

中国版本图书馆 CIP 数据核字 (2008) 第 110093 号

本书本印次封底贴有防伪标。请注意识别。

医学信息搜集的途径与方法

主 编: 聂绍平

出版发行: 人民卫生出版社(中继线 010-67616688)

地 址: 北京市丰台区方庄芳群园 3 区 3 号楼

邮 编: 100078

网 址: <http://www.pmph.com>

E - mail: pmph@pmph.com

购书热线: 010-67605754 010-65264830

印 刷: 北京人卫印刷厂

经 销: 新华书店

开 本: 787×1092 1/16 印张: 39

字 数: 921 千字

版 次: 2008 年 9 月第 1 版 2008 年 9 月第 1 版第 1 次印刷

标准书号: ISBN 978-7-117-10446-3/R·10447

定 价: 73.00 元

版权所有, 侵权必究, 打击盗版举报电话: 010-87613394

(凡属印装质量问题请与本社销售部联系退换)

出版说明

随着医学研究生培养规模的不断壮大,国内研究生培养硬件及软件水平的相对落后与培养高素质研究生之间的矛盾日益突出,如何解决这一矛盾成为我们国家医学研究生培养迫切需要解决的问题。

为了适应新时期国内研究生教育和教学的需要,全国高等医药教材建设研究会、卫生部教材办公室自2004年开始,针对各高校研究生院(处)、研究生导师、在校和毕业后研究生有计划、分期分批地进行了大量、大规模的调研和专家论证工作。在深入探讨“研究生规划教材在研究生培养过程中应该发挥的作用;研究生教材与五年制教材、八年制教材、专科医师培训教材、专著之间的区别与联系”的基础上,根据我国医学研究生教育的实际需要,率先组织策划了这套全国高等学校医学研究生规划教材。

在内容的组织上,该套教材突破传统应试教育教材系统全面的特点,紧扣研究生培养目标,着眼于学生进一步获取知识、挖掘知识和实践创新能力的培养。全套教材包括公共基础课和临床专业课两个系列:公共基础课系列主要围绕研究生科研过程中,从最初的科研设计到最终论文发表的各个环节可能遇到的实际问题展开。临床专业课系列以临床诊疗的回顾·现状·展望为线索,通过对具有转折点意义的诊疗理论、技术或方法探索过程的回顾,目前诊疗中的困惑、局限与不足以及诊疗实践中应注意问题等现状的分析,以及所在学科领域研究热点及发展趋势的展望来探讨新的解决问题的切入点,启发和培养临床创新思维。

该套教材的临床专业课系列主要适用于临床型的硕士生、博士生及相应的临床工作者;公共基础课系列适用于医药卫生各专业的硕士生、博士生及相应的医药卫生工作者。

教材目录

一、公共基础课系列

医学科研课题的设计、申报与实施	主编 李卓娅 龚非力	中英文医学科研论文的撰写与投稿	主编 张学军
医学信息搜集的途径与方法	主编 聂绍平	医学免疫学实验技术	主编 柳忠辉
医学实验技术的原理与选择	主编 李幼平	组织化学与免疫组织化学	主编 李 和 周 莉
医学实验动物学	主编 秦 川	断层解剖学	主编 刘树伟
人类疾病动物模型	主编 施新猷 苏 卫	医学免疫学	主编 曹雪涛
统计分析在医学课题中的应用	主编 蒋知俭	实验室生物安全	主编 叶冬青

二、临床专业课系列

呼吸内科学	主编 钟南山 王 辰	泌尿外科学	主编 杨 勇 李 虹
心血管内科学	主编 胡大一 马长生	妇产科学	主编 曹泽毅
消化内科学	主编 胡品津 刘新光	儿科学	主编 桂永浩 申昆玲
肾内科学	主编 谌贻璞	神经内科学	主编 刘 鸣 谢 鹏
血液内科学	主编 周 晋 黄 河	精神病学	主编 江开达
内分泌内科学	主编 陆召麟 宁 光	眼科学	主编 崔 浩 王宁利
风湿内科学	主编 陈顺乐 邹和健	耳鼻咽喉头颈外科学	主编 孔维佳
普通外科学	主编 赵玉佩 姜洪池	传染病学	主编 李兰娟
骨科学	主编 田 伟 陈安民	急诊医学	主编 黄子通
胸心外科学	主编 胡盛寿	老年医学	主编 张 建 范 利
神经外科学	主编 王忠诚		
血管淋巴外科学	主编 汪忠镐		

前言

随着计算机、现代通讯技术和生物医学信息学的飞速发展,网络上的生物医学资源可谓“浩如烟海”,同时也在与日俱增。网络资源以其独特的时效性、交互性、多样性与经济性,在生物医学信息中的地位也日渐突出,大有部分取代纸质资源的趋势。如何充分利用这些日新月异的网络医学资源,使其充分服务于医学研究与临床实践,促进医学现代化和医学信息社会化,已成为摆在医学工作者(尤其是医学研究人员)面前的实际问题之一。

互联网上医学信息可谓“五彩缤纷”,内容繁多。既有一些较为权威的综合数据库、出版社或期刊站点,又有大量较为自由开放的论坛信息。既有覆盖面广、包罗万象的文摘性数据库,又有实用便捷的全文数据库。既可使用公共搜索引擎,又可使用专业搜索引擎。既能在大型数据库或搜索引擎进行“普遍撒网”式的立体检索,也可直接进入专业数据库进行“重点钓鱼”式的定向检索。既能进行“学习型”的一般性日常检索,又可进行研究型的创新检索。除了传统的馆藏与光盘途径外,还能通过馆际互助、网络求助、电子邮件等多种途径获取所需信息。医学工作者若能有效掌握上述各种医学信息的搜集途径与方法,则能做到足不出户便知天下,事半功倍地掌控学科进展,从而极大地提高临床与科研工作效率。

针对医学文献数据库和网络医学信息资源的变化特点,我们组织部分从事医学研究或临床工作的学者,结合自己的个人实践与认识,编写了这本《医学信息搜集的途径与方法》。本书旨在突破传统概念的文献检索,重点介绍网络医学信息搜集的途径与方法,重在传递检索思想、启迪思路,而不是长篇累牍地罗列一些站点或大量枯燥乏味的信息。本书在形成编写思路和初步编写大纲后,几经征集修改意见而形成最终大纲。

全书共分三篇,基础篇旨在“推本溯源”,重点介绍医学信息的来源与类型等,包括公共与专业搜索引擎、中文和外文全文与文摘型数据库、引文分析数据库、馆藏信息和特种信息等;技巧篇旨在“授人以渔”,介绍一些实用的网络检索策略与技巧,检索策略的编制与调整、全文攻略、创新研究检索等。考虑到临床研究与论文撰写的需要,本篇也系统地介绍了一些医学文献管理的技巧与方法。实战篇旨在“抛砖引玉”,以分子生物学、癌症、药学与循证医学为例,用一些实际的检索个案介绍医学网络信息搜集的技巧与方法。全书共分26章,为提高操作性与实战性,全书采用了大量操作示意图,旨在用轻松的网络语言与电子书风格,将读者带进纷繁而有趣的信息检索殿堂。与国内同类医学信息检索的教材相比,本书强调系统性、新颖性和实用性。全书内容丰富而紧凑,条理清楚,力求全面反映网络环境下医学信息资源检索与利用的新特点、新变化。本书是在深入研究我国高等学校生物医学信息检索教学现状的基础上,结合当前我国信息检索教学的实际和编者多年来的教学经验,吸收国内外生物医学信息检索的最新成果编写而成的。本书既可作为医药卫生专业研究生教材,也可作为临床医生、医学科研和教学人员继续教育的参考用书。

本书的编写工作得到了首都医科大学教材建设经费的资助,得到了首都医科大学附属北京安贞医院研究生处孙海霞、程雯老师的热情支持,得到了汤姆森科技信息集团刘煜总经理、杭州联科美讯生物医药技术有限公司董事长李天天博士的大力支持,在此一并向他们致以最诚挚的谢意。本书的编写工作是在大量现有经验与原创的基础上完成的,也利用了一些相关文献与网络资源,在此也一并向相关作者致以崇高的敬意。

如果本书能达到启迪与开拓信息搜索思路的目的,我们将感到无比的荣幸。尽管本书编写工作历经寒暑,数度易稿,全体作者也倾尽全力,力求完美,怎奈工作繁忙,时间匆促,加之编者水平有限,数字信息资源变化更新较快,书中疏漏和谬误在所难免,恳请广大同仁与读者批评指正。

聂绍平

2008年5月于北京

目录

第一部分 基础篇

第1章	信息检索基本知识	1
第1节	信息检索简史	1
第2节	网络信息的特点、类型与搜索途径	3
第3节	信息检索的基本方法	9
第2章	医学信息资源的主要类型	16
第3章	医学信息资源的搜集途径	28
第4章	通用搜索引擎	37
第5章	医学专业搜索引擎	58
第1节	Health On the Net(HON)	59
第2节	WebMD	65
第3节	HealthAtoZ	73
第4节	Intute	77
第5节	其他医学专业搜索引擎	81
第6章	中文文摘索引型数据库	85
第1节	中国生物医学文献数据库	85
第2节	中国科学引文数据库	88
第3节	中医药数据库检索系统	92
第7章	外文文摘索引型数据库	105
第1节	MEDLINE	105
第2节	PubMed	113
第3节	EMBASE(荷兰医学文摘)	124
第4节	Scirus	132
第5节	Scopus 文摘和索引(A&I)数据库	135
第6节	Biosis Preview(美国生物医学文摘)	140
第7节	CA(化学文摘)	146

第 8 章	光盘数据库简介	154
第 1 节	MEDLINE	154
第 2 节	《中国生物医学文献光盘数据库》.....	158
第 3 节	《中文生物医学期刊数据库》.....	160
第 4 节	中国化学文献数据库.....	161
第 9 章	引文数据库	166
第 1 节	引文索引的起源和历史.....	166
第 2 节	Web of Science	168
第 3 节	中国科学引文数据库.....	186
第 10 章	中文全文数据库	197
第 1 节	CNKI 中国期刊全文数据库	197
第 2 节	维普(VIP)中文科技期刊数据库(Web 版).....	206
第 3 节	万方数据库.....	209
第 4 节	中国医学学术会议论文数据库(CMAC).....	218
第 5 节	超星数字图书馆.....	219
第 11 章	外文全文数据库	221
第 1 节	ScienceDirect 全文数据库	222
第 2 节	SpringerLink 全文数据库	231
第 3 节	Wiley InterScience 全文数据库	238
第 4 节	OVID 数据库平台	246
第 5 节	ProQuest 数据库平台	255
第 6 节	Blackwell Synergy 全文数据库	263
第 7 节	EBSCOhost 数据库平台	268
第 8 节	Ingentaconnect 全文数据库	279
第 9 节	SwetsWise 全文数据库	283
第 10 节	MD Consult 全文数据库	287
第 11 节	Karger 全文数据库	291
第 12 节	IEEE/ET Electronic Library(IEL)全文数据库	295
第 13 节	Netlibrary 电子图书.....	299
第 12 章	出版社与期刊站点	304
第 13 章	馆藏信息查询与利用	312
第 1 节	CALIS 联合目录数据库	313
第 2 节	CALIS 西文期刊目次数据库	319
第 3 节	全国期刊联合目录数据库(联合编目服务系统).....	328
第 4 节	国家图书馆馆藏目录数据库.....	334

第 14 章 特种信息检索	340
---------------------	-----

第二部分 技巧篇

第 15 章 网络检索策略与技巧	355
------------------------	-----

第 1 节 网络检索的基本策略	355
第 2 节 网络信息检索工具的检索功能	357
第 3 节 检索实例	361

第 16 章 检索策略的编制与调整	368
-------------------------	-----

第 1 节 检索词的选择	368
第 2 节 检索式的编制与调整	369

第 17 章 医学全文攻略	374
---------------------	-----

第 1 节 网络文献查找方法	374
第 2 节 馆藏文献查询	381
第 3 节 馆际互借与文献传递	387
第 4 节 网络全文服务	389
第 5 节 向文献作者求助	394
第 6 节 网络文献求助	396

第 18 章 创新研究检索技巧	400
-----------------------	-----

第 1 节 知识框架建立过程中的文献检索	400
第 2 节 Web of Science 辅助获得研究设想	404
第 3 节 研究课题确立过程中的信息获取	407

第 19 章 论坛信息的综合利用	409
------------------------	-----

第 20 章 医学文献管理软件的应用	417
--------------------------	-----

第 1 节 参考文献及其管理软件概述	417
第 2 节 EndNote	419
第 3 节 NoteExpress	437

第三部分 实战篇

第 21 章 分子生物学信息数据库	443
-------------------------	-----

第 22 章 癌症信息数据库	462
----------------------	-----

第 23 章 药物领域相关数据库	477
------------------------	-----

第 1 节 药物研发生命周期各阶段信息需求分析	477
第 2 节 数据库具体应用实例演示	483

第 24 章 临床医学网络资源及利用	533
第 25 章 循证医学专业数据库	547
第 26 章 引文检索和分析实战篇	571
第 1 节 Web of Science 检索分析实例	571
第 2 节 中国科学引文数据库检索分析实例	607
主要参考文献	613

第一部分

基础篇

第 1 章 信息检索基本知识

信息检索 (Information Retrieval) 通常指文本信息检索,是将信息按一定的方式组织和存储起来,并根据信息用户的需要找出有关信息的过程,包括信息的存储、组织、表现、查询、存取等各个方面,其核心为文本信息的索引和检索,所以其全称又为“信息的存储与检索 (Information Storage and Retrieval)”。从历史上看,信息检索经历了手工检索、计算机检索到目前网络化、智能化检索等多个发展阶段。检索的内容也从相对封闭、稳定、独立的数据库扩展到开放、动态、更新快、多元化、分布广泛、管理松散的网络内容。信息检索的用户也由原来的专业情报人员扩展到普通大众。随着 Internet 的飞速发展和网络信息资源的激增,网络信息检索已成为科研人员与医学工作者必备的基本技能之一。

第 1 节 信息检索简史

信息检索作为一门学科,其历史可追溯到 20 世纪中期。在此以前,信息存储和传播主要以纸质介质为载体,信息检索活动也围绕着文献的获取和控制展开。随着社会与科学的进步,信息检索经历了从手工检索、机械检索、计算机检索到网络智能化检索的发展过程。

一、手工检索时代

在互联网出现或普及之前,手工检索是信息检索的主要方式。手工检索 (manual retrieval) 是以手工翻检的方式,利用工具书 (包括图书、期刊、目录、卡片等) 来检索信息的一种检索手段。19 世纪下半叶,美国的公共图书馆和大专院校图书馆开始提供正规的参考咨询工作。20 世纪初,多数图书馆成立了参考咨询部门,主要利用图书馆的书目工具来帮助读者查找图书、期刊或现成答案。随着文献的激增和读者需求的增长,逐渐发展到从多种文献源中查找、分析、评价和重新组织情报资料,“索引”突破了以前的狭隘范畴,成为独立的

检索工具。到 20 世纪 40 年代,又进一步包括回答事实性咨询,编制书目、文摘,进行专题文献检索,提供文献代译等。“检索”从此成为一项独立的用户服务工作,并逐渐从单纯的经验工作向科学化方向发展。

手工检索无需特殊设备,用户根据所检索的对象,利用相关的检索工具就可进行,其方法比较简单、灵活,容易掌握。但是,手工检索耗时、费力,特别是进行专题检索和回溯性检索时,需要翻检大量的检索工具反复查询,花费大量的人力和时间,而且很容易造成误检和漏检。

二、脱机批处理检索

50 年代中期至 60 年代中后期是信息检索的脱机批处理阶段。当时计算机还没有连接通信网,也没有远程终端装置,不能提供实时检索,只能进行现刊文献的定题检索(Selective dissemination of information)和回溯性检索(Retrospective search),或利用计算机编辑出版检索性刊物。1954 年,美国海军机械实验中心使用 IBM701 型机,初步建成了计算机情报检索系统,这预示着以计算机检索系统为代表的信息检索自动化时期的到来。

三、计算机联机检索

1965 年美国系统发展公司(SDC)研制成功 ORBIT 联机情报检索软件,开始了联机情报检索系统阶段。与此同时,美国洛克希德公司研制成功了 Dialog 检索系统。随着卫星通讯技术、微型计算机以及数据库生产的同步发展,到 20 世纪 70 年代,使用户得以冲破时间和空间的障碍,实现了国际联机检索。远程实时检索多种数据库是联机检索主要的优点。计算机检索技术从脱机阶段进入联机信息检索时期。联机检索是计算机技术、信息处理技术和现代通信技术三者的有机结合。

四、网络化联机检索

由于单纯的手工检索和机械检索都存在一定的缺陷,亟需新型的信息检索方式。20 世纪 60~70 年代,随着因特网、通信和网络传输技术的发展,信息检索突破了发展瓶颈。

因特网(或互联网)是计算机交互网络的简称(Internet),最早起源于美国国防部高级研究计划署 DARPA(Defence Advanced Research Projects Agency)的前身 ARPAnet,于 1969 年投入使用,成为现代计算机网络诞生的标志。互联网的发展大大促进了信息搜集和检索技术的发展和运用,一大批优秀的搜索引擎和专业数据库,为我们提供了很好的快速获取信息的方法。互联网信息检索所具有的多样性、灵活性远远超出了传统的信息检索,我们需要继承与沿用在传统信息检索中已经形成的某些检索思维模式以及一些已成定式的检索方法,更需要掌握互联网信息检索具有的特点、了解影响信息检索的因素,通过实践提高获取信息的能力。

因特网犹如一个浩瀚的信息海洋,信息检索首先应解决的问题就是通过什么途径去访问它,从而检索到自己所需的信息资料。传统的信息访问手段主要有远程登录(Telnet)和文件传输(FTP)两种,两者要求使用者不但应了解所需信息在网络中的存放位置,还必须掌握有关的操作命令。为减轻用户负担,近年来,一些方便易用、智能化的信息检索工具相继

问世,主要有 Gopher、WAIS、Archie 和 WWW 等,其中以 WWW 应用最为普及。Gopher 是因特网提供的一种菜单式驱动信息检索工具,因特网上的 Gopher 服务器将信息资源组织成单一形式的资料库,称作 Gopher 空间。Gopher 以关键字作索引,采用客户机/服务器工作模式。使用 Gopher 有两种方式:一是使用本地机的 Gopher 客户程序,二是通过 Telnet 登录到安装有 Gopher 客户程序的机器上使用 Gopher。WAIS(Wide Area Information Service)译为“广域信息服务”,是检索巨大信息空间的快速而有效的工具。用户使用 WAIS 时,与 WAIS 的客户程序打交道,访问检索由 WAIS 服务程序完成。用户可通过安装了 WAIS 客户程序的本地机或登录到提供公众 WAIS 程序的机器上,对公布在因特网上不同服务器中的各种数据库进行检索。目前,因特网上有数百个免费的 WAIS 资源,涉及生物、天文、地理、农业、人文等各类信息。Archie 是加拿大麦吉尔大学设计的一种专门查询文档的工具软件,配有这种软件的文件寻找服务器会定期自动访问网上数以万计的服务器,将这些服务器上的文件形成可检索的数据库,从而为用户提供文件出处服务。除上述网络信息检索工具外,还可利用 BBS、Usenet 查找网络动态性新闻、专题讨论等信息。

网络信息检索系统包括了计算机在信息检索领域表现出来的全部优点。它是联机检索的高级阶段,使人们可以在很短的时间里查找到全球的信息。值得指出的是,网络信息环境的出现,使得信息检索研究的对象和范围不断扩大,研究队伍也突破了原有的以图书情报领域的专家学者为主的框架,众多的信息公司加入到研究开发信息检索系统的行列。可以说,网络使计算机信息检索技术进入一个崭新发展阶段,而网络信息检索又使得网上信息源利用率提高,信息组织更为有序和高效。总之,基于因特网的检索系统成为网络信息检索系统的代表。

第 2 节 网络信息的特点、类型与搜索途径

随着 Internet 的飞速发展,网络信息日渐成为信息检索的主要来源。由于电子信息资源存在内容丰富、更新及时、便于存储和查阅等优势,大有逐步取代传统纸质资源的趋势。

一、网络信息的特点

网络信息资源是以电子计算机技术、通信技术、多媒体技术相互融合而形成的以电子网络为传输性载体和传输媒介的信息资源,是一种新型数字化资源,与传统文献相比有较大的差别。了解网络信息资源的特点,有助于用户对其利用。从信息检索的角度来讲,网络信息资源具有以下特点:

1. 内容丰富,信息量大 网络信息资源极为丰富,因特网已经成为继电视、广播和报纸之后的第四媒体,几乎是无所不包,且类型丰富多样,如学术信息、商业信息、政府信息、个人信息等。它一方面给用户选择提供了较大余地;但另一方面,大量的毫无价值的冗余信息也给用户造成了很大的麻烦。

2. 信息层次多,品种多样 因特网上的信息资源层次众多,有一次信息、二次信息、三次信息;有文本信息、图像信息、图形信息、表格信息、超文本信息等。包括各种电子书刊、书

目数据库、联机数据库、软件资源等,是多媒体、多语种、多类型信息的混合体。网络信息资源包罗万象,几乎覆盖各学科、各领域、各地域、各语言的信息资源;信息发布者既有政府部门、高等院校、科研院所、学术团体、行业协会,更有大量的公司企业和个人。

3. 自由发布,交流直接 网络信息资源中包括很多非正式交流渠道发布的信息,包括那些正式出版物中所不能得到的信息,如类似于灰色文献的信息、还未成熟的观点、个人网页上的研究心得、教学资料等。同时,因特网扩大了人际交流的空间,如新闻组、讨论组、邮件列表等,都为用户提供了更多的直接交流的机会。

4. 信息传播快、更新及时 与传统文献相比,网络信息变化快、可随时发布,更新及时,具有活动性,可根据需要不断扩充。其信息传播速度十分快,还能实现即时信息交流。

5. 信息资源多媒体化 近年来,多媒体声像网络信息呈日渐增多的趋势,有的支持在线视听,有些还能提供音频或视频下载,非常直观。部分还能提供 Podcast 格式,能下载到 ipod 播放器直接使用,十分便捷。

6. 关联度强,检索快捷 网络信息资源利用超文本链接,构成了立体网状文献链,把不同国家、不同地区、各种服务器、各种网页、各种不同文献通过结点链接起来,关联度增强了。通过各种专用搜索引擎及检索系统使信息检索也变得方便快捷。

7. 分散无序、缺乏管理 由于任何机构、个人都可自由地在网上发布信息,信息的发布有很大的自由度和随意性。没有统一的标准,缺乏必要的过滤、质量控制和管理机制。而且,“海量”的信息和快捷地传播加剧了网络信息的无序状态,许多信息资源缺乏加工和整理,只是时间序列的信息堆积,缺乏系统性和组织性。因特网的信息内容十分庞杂,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、色情等不健康信息混为一体,既有大量国际水平的研究成果,又有许多难登大雅之堂的涂鸦之作和虚假信息,信息质量良莠不分、参差不齐,给网络信息的利用带来极大的不便。这也给使用者增加了一定难度。

二、网络信息的类型

网络信息资源数量庞大,类型繁杂,形式多样,广泛分布在网络之中,没有统一的组织管理机构,也没有统一的目录。对网络信息资源进行分类,有助于深入地了解网络信息资源。

1. 按发布形式划分 按发布形式划分,可将因特网上的信息资源分为以下类型:①书目、索引、文摘型的二次文献数据库:如 ISI 网站上的“Web of Science”是著名的“SCI”、“SSCI”、“AHCI”的 Web 版,能查到各类引文数据;②查阅知识条目的参考型三次文献数据库:如美国 Gale 网站上的“Gale Net”专栏,能查阅该出版社编辑出版的各类参考工具书;Encyclopedia Britannica 网站和 Webster's Dictionary 网站分别是《不列颠百科全书》和《韦氏英语词典》的 Web 版;③期刊文献数据库:如荷兰 Elsevier Science、SwetsNet、Kluwer、OVID,美国 ProQuest UMI、Ebsco、IEEE、Wiley,德国 Springer-link,瑞士 Karger,英国 Ingenta、Blackwell-synergy,以及我国的 CNKI 等,在其网站上可查到其收录期刊全文数据库;④其他文献型数据库,如英国的 Derwent 网站可查阅 Derwent 专利全文数据库;美国的 UMI 网站可查阅学位论文数据库;我国的知网 CNKI 优秀硕博学位论文检索库等。

2. 按信息交流的方式划分 可分为:①非正式出版信息:指流动性、随意性较强的,信息量大、信息质量难以保证和控制的动态性信息。如电子邮件、专题讨论小组和论坛、电子会议、电子布告板新闻等工具上的信息;②半正式出版信息:又称“灰色”信息,指受到一定产权保护但没有纳入正式出版信息系统中的信息。如各种学术团体和教育机构、企业和商业部门、国际组织和政府机构、行业协会等单位介绍宣传自己或其产品的描述性信息;③正式出版信息:指受到一定的产权保护,信息质量可靠、利用率较高的知识性、分析性信息,用户一般可通过 Web 查询到。如各种网络数据库、联机杂志和电子杂志、电子图书、电子报纸等。它们或是传统出版物的数字化,或是有明确创建者、有版权的直接网络出版物。

3. 按照信息加工程度划分 可分为:①一次网络信息资源:指因特网的原始信息,包括电子图书、电子期刊、电子报纸等;②二次网络信息资源:对一次网络信息资源的搜集、加工和处理,主要指搜索引擎、虚拟图书馆等,是网络检索工具的重要组成部分。这类网络信息资源是用户经常利用的工具,是获取一次网络信息资源的门户和入口;③三次网络信息资源:对二次网络信息资源的搜集和对已搜集二次网络信息资源的组织,以元搜索引擎为其代表(下述)。

4. 按照工作方式划分 可分为全文搜索引擎(Full text search engine)、目录索引类搜索引擎(Search index/directory)和元搜索引擎(Meta search engine)。全文搜索引擎是名副其实的搜索引擎,国外具有代表性的如 Google、百度等。目录索引虽然有搜索功能,但在严格意义上算不上是真正的搜索引擎,仅仅是按目录分类的网站链接列表而已。用户完全可以不用关键词查询,仅靠分类目录也可找到需要的信息。最具代表性的有 Yahoo、Open Directory Project(DMOZ)、LookSmart、About 等。国内的搜狐、新浪、网易搜索也都属于这一类。元搜索引擎在接受用户查询请求时,同时在其他多个引擎上进行搜索,并将结果返回给用户,使搜索结果更科学更全面。著名的元搜索引擎有 InfoSpace、Dogpile、Vivisimo 等。

5. 按发布机构划分 因特网上的信息资源是各机构发布信息的总和,由于发布机构的特点和性质不同,其信息类型也各有特点。根据发布机构可将网络信息资源分为:①企业站点信息资源:一般以 COM 为一级或二级域名注册。如 www.medtronic.com、www.cordis.com 等,能提供公司产品规格、认证、研究等多方面信息;②学校、科研院所站点信息资源:一般以 edu 或国家域名为一级或二级域名注册。如哈佛大学 www.harvard.edu、加拿大多伦多大学 www.utoronto.ca 等,信息表现为初始源和再生源共存,主要提供学术性较强的各种信息,如科研活动介绍、学术动态、信息检索、远程教育等;③信息服务机构站点信息资源:一般以 net、com、gov,或行政区域为一级或二级域名注册,如:<http://www.cnnic.net.cn>(中国互联网络信息中心)等。这类站点一般以再生源信息为主,开展信息资源的开发与利用服务、网络功能的开发与应用服务,如全文数据库查询、建立搜索引擎、代理域名注册等;④行业机构站点信息资源:一般以所属上级部门为域名注册,有 com、ac、gov 等,如 www.cardiosource.com、中国医师协会 www.cmda.gov.cn 等。

此外,根据传输协议,还可将网络信息分为基于超文本传输协议(HTTP)的信息资源、基于文件传输协议(FTP)的信息资源、基于远程登录(Telnet)的信息资源、用户服务组信息资源(如新闻组、电子邮件群、邮件列表、专题讨论组)和 Gopher 信息资源等。

三、网络信息的搜索途径

(一) 生物医学数据库

1. 文摘型数据库 能提供原文的摘要信息,包括题目、作者、刊名、摘要以及与论文相关的信息等。部分数据库还同时提供部分文章的全文链接(如 Pubmed 的 Linkout)。该型数据库的特点是收录数据量大,收录文献类型、出版形式和语种较多,收录文献的回溯时间较长,因此成为信息检索的主要工具之一。在生命科学领域中,主要文摘型数据库包括美国全科医学文献索引(Medline)、荷兰医学文摘(Embase)、美国生物医学文摘(Biosis Previews)、美国化学文摘(CA on CD, SciFinder)等。文摘型数据库有多种操作平台,例如,Medline 数据库几乎囊括了所有生物医学领域,其应用平台包括光盘数据库、网络版数据库和国际联机数据库等,以网络版数据库最为常用。目前,有多个专业站点提供网络 Medline 服务,如 Pubmed、Medline@OVID、Medline@EBSCO、Medline@Web of Knowledge、Medline@CSA 等。

2. 全文数据库 互联网技术的迅猛发展使得期刊电子化成为可能。目前,生物医学领域的主要期刊已基本实现电子版与纸质版同时发行。除少数期刊仅仅在自身站点提供全文服务外,多数期刊的全文服务被一个或多个全文数据库整合。例如, *Circulation* 除在 AHA 站点(<http://circ.ahajournals.org/>)提供全文服务外,还被 OVID、EBSCO 等全文数据库收录。由于多数全文数据库既能提供文摘式搜索(Search)服务(Powered by Medline),若权限许可,又能同时下载全文,因而其使用人数已远远超过文摘型数据库。

3. 事实型数据库 是一种存放某种具体事实、知识数据的信息集合。在对现实世界进行定性和定量描述或研究时,往往需要事实情报,如化合物的结构式或分子式,物质和材料的特性,指数,设备型号,地名等。事实(fact)一词来源于拉丁文 factum,意思为已做成的,已完成的。为满足这一要求,把某一学科已知的事实数据搜集起来建成数据库,就是事实数据库。如数值数据库、指南数据库、术语数据库等,相当于印刷型文献中的字典、辞典、手册、年鉴、百科全书、组织机构指南、人名录、公式与数表、图册(集)等。如:Beilstein、ISI Chemistry、LEXIS.COM、GaleNet、中国资讯行数据库、科学数据库系统、人类生物基因组数据库等。利用事实数据库可以查找已知事实,或判断未知事实。

4. 引文分析数据库 以美国科学情报研究所(Institute for Scientific Information, 简称 ISI, 网址:<http://www.isinet.com/>)出版的科学引文索引数据库最为著名。其出版形式包括印刷版期刊和光盘版及联机数据库,现在还发行了互联网上 Web 版数据库。其中以生命科学及医学、化学、物理所占比例最大,收录范围是当年国际上的重要期刊,尤其是它的引文索引表现出独特的科学参考价值,在学术界占有重要地位。

(二) 搜索引擎

搜索引擎是专门提供信息查询服务的网站安装在其服务器上的一种能在网上漫游并搜集它所得到的信息,同时自动生成本地索引的软件。它是因特网上最为有效的信息导航工具,是网上冲浪的得力助手。搜索引擎并不是真正搜索互联网,它实际搜索的是预先整理好的网页索引数据库。通常意义上的搜索引擎指收集并索引了互联网上几千万到上百亿不等的网页,建立索引数据库的全文搜索引擎。由于这种服务都是一些著名的站点提供的,所以