

●王云娣

基于网络的社科信息资源分布及检索策略研究^{*}

摘 要 网络社科信息资源包括:电子期刊资源、图书信息、报纸资源、学位论文资源、电子论坛和电子会议、网络数据库等。可直接进入网站或数据库、利用搜索引擎、利用专业信息资源导航系统获取信息资源。参考文献 6。

关键词 网络 社科信息资源 检索策略

分类号 G350

ABSTRACT Network-based social science information sources include electronic journals, book information, newspaper resources, dissertations and theses, electronic forums and electronic conferences, and online databases. We can access these sources by direct access, search engines and special subject navigation systems. 6 refs.

KEY WORDS Network. Social science information resources. Search strategy.

CLASS NUMBER G350

随着全球信息化的不断深入,网上社科信息资源日益丰富。但由于网上信息资源高度分散、数字信息的易拷贝等特点导致信息重复率高,无序性更为突出。所发布的社科信息没有统一的管理和规范,加之现有的检索工具智能化程度较低,导致信息查询困难,大大影响了网上社科信息资源的利用效率。了解目前网上社科信息资源的分布情况,加快和扩大网上信息资源开发利用的进度和规模,利用信息网络共享社科信息资源,弥补目前普遍面临的信息资源短缺的问题,已经迫在眉睫。

1 网络社科信息资源的分布

1.1 网上电子期刊资源

网上电子期刊具有时效性强、内容丰富、检索途径多样等优势,利用率非常高。网上电子期刊包括与印刷版同时发行和仅在网上传行的两种,其主要来源渠道有:

(1) 出版商和文摘索引服务商提供的期刊检索服务。目前,国际上大多数主要的期刊出版商都建立了全文数字化期刊数据库,并通过因特网提供服务。如美国学术出版社(Academic Press)的 IDEAL(International Digital Electronic Access Library)系统(<http://www.idealibrary.com>),德国 Springer-Verlag 出版公司的 Link 系统(<http://www.link.springer>。

de)等。有的文摘索引服务商则将自己的产品上网,提供包括期刊订购、检索、全文传递等服务。如美国威尔逊公司(Wilson)的 Wilson Fulltext(<http://www.wilson.com>)。

(2) 网络数据库信息服务商提供的服务。存储在网络服务器上的网络数据库品种多、质量高,大多由信息服务商或大型图书情报机构创建维护,如世界上最早和最大的国际联机检索系统 Dialog,美国联机计算机图书馆中心 OCLC,中国期刊网等。访问的形式主要有授权直接访问或访问镜像站上的数据库两种。中国期刊网的题录数据库免费提供用户检索,可为用户提供 6000 多种刊物的题录信息。有的数据库服务商不定期地提供某些数据库的免费试用服务。

(3) 文献情报部门或学术性机构提供的服务。文献情报部门或学术性机构依托自己或协作单位的馆藏,通过自己的网站提供期刊或其他类型文献的网上检索服务,用户在有的网站检索二次文献信息的同时,还可以订购原始文献,原文可通过 E-mail、传真、浏览、下载和联机打印等方式获取。如美国科技信息研究所的科学网(web of science, <http://wos.isiglobalnet.com>)提供世界三大引文索引 SCI、SSCI、A &HCI 的检索。

(4) 网络版期刊提供的服务。许多期刊正在网上建立自己的网站,便于读者上网检索或提供免费

^{*} 本文系浙江省教育厅 2001 年资助项目的研究成果。

的电子邮件现刊目录服务。如《中国远程教育》免费为广大读者提供全文阅读、期刊订阅及检索服务;《世界华人名人录》网络版提供网络版的题录和全文;中国经济展望网提供《宏观经济观察》、《宏观经济论坛》等多种网络版期刊的题录和全文,等等。

1.2 网上图书信息

网上提供的图书信息主要包括书目信息和电子图书。

(1) 出版商提供的书目信息。国内外许多出版社都设有自己的网站,通过因特网发布其出版物的目录、最新图书内容简介、价格等信息。如高等教育出版社提供新书推荐、教材目录以及书目查询服务等;Academic Press 提供其出版物的目录、最新图书内容简介、价格等信息,带有一定的销售宣传性质。

(2) 图书馆、文献情报中心提供的联机图书馆馆藏目录。目前全球至少有 15000 个大型图书馆的联机目录 24 小时对外开放。读者通过 OPAC 可以查询到大量的书目信息,确定某本书是否在馆,能否通过馆际互借加以利用等。有些联机目录不但提供书目及摘要,还提供其他信息资源,用户可以通过相应的 URL 地址访问世界各个国家和地区的各种图书馆。

(3) 数字图书馆提供的电子图书。数字图书馆是数字化生存时代的重要文化基础设施。用户只需通过鼠标和键盘,记载人类文明成果的图书资料就会方便地进入人们的生活。如超星数字图书馆可为读者提供 35 万册的 PDG 格式的数字图书。其会员可随时阅览到最新图书。国内可供读者阅读图书的还有中国数字图书馆、书生之家数字图书馆等。

(4) 研究机构提供的电子图书。以香港中文大学中国文化研究所的“古文献资料库中心”和中国台湾的“瀚典”为代表。两网站除部分内容需付费使用外,绝大部分内容可免费在线浏览。

(5) 个人或团体网站提供的电子图书。大多数文学类网站属于此类,有的以收藏为主,有的以链接为主。虽然侧重点有所不同,但大多是中外文学名著和眼下一些比较红火的图书,也包括一些网友自己的习作和文摘,一般为全部免费提供给读者阅读。

1.3 网上报纸资源

网上电子报纸时效性强、传播范围广,提供大量时事、专题、经济、政治、法律法规等方面的信息,正

逐渐成为新的重要的社科信息源。据不完全统计,自 1995 年 10 月《中国贸易报》率先在网上开办电子版以来,到 1999 年 6 月,国内已有 273 家报纸上网提供阅览。以《人民日报》网络版为例,它从每天凌晨 0 时 30 分到 5 时,将每天出版的《人民日报》、《市场报》等 17 种报刊的全部文字和图片上网发布,还分为“中国”、“国际”、“财经”等 9 个频道,每天滚动发布国内外新闻 1500 条,并就新闻热点制作专题。网站还设置了对报刊和频道新闻均有效的信息检索功能,读者很方便就能查到相关资料。

1.4 网上学位论文资源

(1) 利用全国性的学位论文数据库提供的信息。许多国家已将学位论文数字化列为重要的项目计划,其中较具规模与成效的,例如:ETD Digital Library Networked Digital Library of Theses and Dissertations (NDLTD) 网络博硕士论文数字化图书馆计划。它是一个国际论文资源的电子图书馆计划。通过整合性检索可检索 19 个单位的论文;另外更列出许多大学及研究机构博硕士论文的网址。CALIS 高校学位论文(文摘)数据库,是中国高等教育文献保障系统的一个子项目,建库的目的是通过对分散在各大大学中各类学科的学位论文文献信息的收集、整理、建库、上网,使原始的论文信息获得升值,为国内外希望获取高校学术信息的用户提供方便的查询途径,起到推动高校教学、科研交流和促进发展的作用。高校学位论文数据库收录包括北京大学、清华大学等全国著名大学在内的 83 个 CALIS 成员馆的硕士、博士学位论文,截止到 2002 年 5 月 21 日,已有 97 所学校申请加入学位论文数据库建库工作,学位论文数据库上网数据已达到约 9.7 万条。

(2) 利用商业出版机构或信息中介者提供的信息。目前已有商业出版机构取得授权,通过正式的出版渠道出版学位论文;有些则是委托代理商接受论文订购。如 1983 年成立的 UMI 公司,是美国学术界首屈一指的出版公司,出版物包括自 1861 年美国的第一本论文到目前所出版的 150 多万本美加地区博硕士论文,1997 年以后出版的已经有 PDF 形式的论文,可直接使用信用卡订购下载。通过 Amazon 网络书店订购学位论文,前 25 页可免费浏览。

(3) 利用联机检索系统中的学位论文数据库。一些知名的联机检索系统中都包含大量的与学位论文有关的数据库。如 Dialog 系统、Questel Orbit 系统、STN FizKar/sruhe 系统。我国的“中国优秀博硕

士学位论文全文数据库”由教育部主管、清华大学主办,是中国知识基础工程(CNKI)继《中国学术期刊全文数据库》等超大型数据库建成后,又一项大规模开发利用我国知识信息资源的重大信息化工程。按照CNKI信息采集范围规定的信息源,它已收录全国有博士招生点的280多家院校的1800多个博士点和3000多个硕士点2000年以来的博硕士学位论文5万多本。并按科学的体系进行了编辑。到2002年底,加入这一数据库的博硕士培养单位将达到80%以上,2002年6月2日的《光明日报》称该数据库是目前世界上惟一最科学、最全面的中文全文博硕士学位论文检索数据库。

(4)学术教育单位自建的学位论文数据库。有些大学或资料中心在原有图书馆馆藏目录查询系统之外建立本校的博硕士论文查询系统,部分学校除提供书目及摘要外,另有全文下载功能。如:MIT Theses Online 麻省理工学院,收录经选择后的博硕士论文,提供论文扫描后之全文。可线上订购全文,包括:纸本、PDF档及微缩片三种形式。《上海交通大学学位论文数据库》,提供多种途径包括摘要和全文的检索能力,可以查到上海交通大学历年的博硕士论文书目资料及摘要,博士论文的全文下载仅限授权用户。

1.5 网上政府信息

政府信息是政府活动的原始记录和产物,它的发展与政府机构本身的扩充及其职能的强化有着不可分割的联系;其文献地位和使用价值也随着综合国力的增强和在国际事务中所发挥的作用,而不断上升和升值。网络政府信息的多少,已被视为一个国家民主程度的表征之一。

在网络发达的美国,据1997年中期统计,美国联邦政府的4300个网站和215个电脑公报牌已将政府的42个部门机构控链为一体,在2000年末已将政府3/4的议事录(transaction)电子化。通过网络,可以查找美国的总统文献,如:国情咨文、经济咨文等;可以通过美国GPO Access主页、美国国务院网站或美国国务院外交事务网等查找美国国务院的文献资料;还可通过美国商业部的国情普查局、劳动统计局、经济分析局等查找美国的统计资料。

1999年1月22日,中国电信和国家经贸委经济信息中心联合40多家部委信息主管部门在北京召开了“政府上网工程”启动大会,要求1999年底60%以上的部委和政府部门在163、169网上建立正式网站,到2000年底则要实现80%以上部委和政府部门

建立网站。据2002年7月22日中国互联网络信息中心(CNNIC)第10次中国互联网络发展状况统计报告,全国各级政府部门申请gov.cn域名的已达5864个,“政府上网”已经成为现实。

1.6 电子论坛和电子会议

因特网上设有USENET及Listserv电子论坛,也称新闻讨论小组。USENET及Listserv都是由成千上万个专题讨论小组构成。每个小组是由某一主题参与的文章所构成。USENET与Listserv类似,但是,前者范围更广泛,几乎无所不包,一般不需订购便可参与;而后者较为严肃,而且更趋学术性,通常还需订购方可参与。因特网上用户通过E-mail均可自由参与电子论坛的活动,从中可以获得任何其他手段都难获得的第一手重要专题信息与资料。因此,它是研究人员及时了解跟踪学科动态与前沿的最有效途径之一,可消除印刷出版物时间滞后的缺点。电子论坛的另一特点是举办国际电子会议。例如,利用Listserv,一些学术团体与组织已成功举办了多次专业性的国际学术会议。参加这些会议,能了解本专业的最新研究发展动态,获取完整的会议论文与资料。

1.7 网络数据库

存储在网络服务器上的文献信息数据库称为网络数据库,多由信息服务商或大型图书情报机构创建维护,品种多,质量高,但基本上都是需要付费才能使用。访问网络数据库的形式主要有直接访问和访问镜像上的数据库两种。目前在网络上运行的著名数据库,如:Dialog系统已拥有600多个集文献信息库、数据信息库、事实信息库和全文信息库于一体的大型专业数据库系统;OCLC Firstsearch数据库,能为用户提供包括Article First、Contents First、ERIC、Fast Doc、GPO等14个基础数据库。

1.8 网上其他社科信息

网络上还有大量的会议信息、政策法规、标准文献、产品样本目录、组织机构、通讯讨论组等。这些资源的实用性也很强。

2 网络社科信息资源的检索

2.1 直接进入网站或数据库进行检索

通过这种途径进行检索,要求检索者对网络信息资源分布有所了解,进行检索前,首先要分析所检课题的信息类型,然后找到可能包含其内容的相应网站的相应系统。比如,要了解某一新课题相关论著,首先确定需要检索书目,可以利用国家图书馆的

书目检索系统进行检索,也可进入高校图书馆书目信息查询系统进行检索。如果要了解课题目前国内外研究最新进展,首先确定可以检索期刊目录,外文的可以进入清华的科技新刊报导检索,中文的可以检索 CNKI 的期刊专题题录数据库。

2.2 利用搜索引擎进行检索

搜索引擎俗称“网络门户”,是搜集网络信息资源最方便的入手工具。据统计,至少 85% 的网上用户使用搜索引擎来查找信息。目前在因特网上已有数千个搜索引擎,中文搜索引擎也有 100 多种。利用搜索引擎可以搜索所有在网上的资源而不必预先知道资源所在的地址。搜索引擎通常有两种检索方式:一种是分类目录导航,一种是输入检索词进行匹配检索。前一种是以 Yahoo! 为代表的人工分类式搜索引擎,根据检索主题所属类别,逐级展开分类目录,查找分类匹配信息。分类式搜索引擎虽具有分类清晰、准确的优点,但却存在人工成本高、收集内容有限、难于更新的不足。后一种是以 AlotVista 为代表的基于 Robot 的搜索引擎,它提供对关键词、主题词或自然语言的查询,由程序自动搜索。用户只要在搜索框中输入检索式或表达式,搜索引擎就会返回一组指向相关站点的超链接。由于是机器人程序自动搜索,可方便地收集更多的网站并及时更新、发现及删除已不存在的站点,从而大大提高用户查询结果的数量。这两种类型搜索引擎的工作原理各不相同,检索效果也各有差异,应合理选用。

2.3 利用专业信息资源导航系统获取信息资源

搜索引擎虽然是整合互联网信息资源,并使之有序化的重要手段。但是据 NEC 研究院的最新研究,网上信息缺乏有效整合标准和手段,搜索引擎只能覆盖网上不到 1/4 的信息,而发表在《自然》杂志的研究称,没有一个搜索引擎可以覆盖网上信息内容达 16% 以上,并且随着站点越来越多,这个比例还会下降。因此通过通用搜索引擎查找信息具有一定的局限性,因为通用搜索引擎的信息覆盖面广,面向大众的信息居多,要查询业务性较强的主题,查准率较低。为了使网上的信息资源充分发挥作用,许多国家的信息工作者把分散在网上的某一或某些主题信息通过多个搜索引擎的协同搜索,选择有较高利用价值的信息资源 URL,经过专业人员的加工、总结、

归类,以目录的形式组织起来,按学科门类进行分类,使用户能按类浏览,方便地获得某一主题的相关信息。通常先按照学科设立一级类目,在一级学科类目下设二级学科栏目,二级学科类目下可以按照信息资源类型设三级类目(如果本学科的信息资源较少,可以不再按信息资源类型设三级类目,使二级类目直接对应于具体的信息资源,这样更简洁明了),三级类目对应的就是具体的信息资源的网址。从而形成了一系列的网上专业资源指南和 www 虚拟图书馆,把读者和网上信息资源连接起来,可以节省用户网上漫游的时间。如:知识管理虚拟图书馆(<http://www.brint.com/km/>)、语言类虚拟图书馆(<http://www.june29.com/>)等。

3 基于网络的社科信息资源检索策略

网上社科信息资源内容丰富、种类繁多、更新频繁,通过各种专业搜索引擎和专业信息资源导航系统查找是最主要的途径。但是,任何一个研究人员都必须关注、浏览本专业相关的国际学会、协会、研究所的网上主页,同时加以取舍和鉴别,最终从网上获得有价值的信息,这才是网络信息检索的主旨所在。同时,网上的信息检索系统有着各自不同的服务功用,用户在检索时,如果选用不恰当,会事倍功半,因此在实际检索中应针对不同的检索要求,加强检索策略的规划,其中包括检索标识的确定、检索入口的选择、检索词间的逻辑关系以及检索提问的表达等。正确而完善的检索策略可优化检索结果,反之会造成检索的失误或失败。

3.1 确定检索项

(1) 分析检索课题,明确检索要求。分析检索课题,是为了弄清楚检索课题要达到的目的和意图,弄清楚课题学科属性、专业范围及相关内容,以便正确选用数据库、检索词和逻辑运算符,确定检索标识,达到良好的检索效果。

(2) 选择适当的语词,确定检索项。规范词和自由词是检索常用的语词。规范词是规范化的语词,是受控词汇,它取自主题词表(叙词表)。自由语词是自然状态的语词,属于非受控词汇。检索时,应遵循优先使用规范词,根据实际情况恰当使用自由词的原则。

●翟晓娟

基于 php 的电子期刊检索网站建设核心技术

摘要 电子期刊检索网站的开发环境,包括操作系统、Web 服务器和开发工具。Windows2000 + Apache + PHP 构架了电子期刊检索网站的开发平台。参考文献 3。

关键词 因特网 网站建设 电子期刊检索

分类号 G250.76

ABSTRACT The environment for the development of websites of electronic journal retrieval includes operation system, Web server and development tools. In this paper, the author discusses a platform based on Windows2000 + Apache + PHP. 3 refs.

KEY WORDS Internet. Website development. Electronic journal retrieval.

CLASS NUMBER G250.76

期刊资源信息含量大,出版周期短,传播速度快,是广大读者首选并被广泛利用的最有价值的信息源。随着信息技术的飞速发展以及网络时代的到来,期刊资源更是日益丰富,种类不断增多。传统的

图书馆期刊资源建设理念已经不能适应形势的变化。因此,基于因特网的电子期刊服务引起图书馆人的巨大关注。建设电子期刊检索网站,可以使期刊充分发挥学术价值、信息价值和参考价值,更好地

(3) 检索项之间进行逻辑组配,编制检索表达式。不管是数据库还是网上的搜索引擎,几乎都具有逻辑组配检索功能。由于已确定的检索项表达的主题概念间存在逻辑关系,可以通过使用布尔逻辑算符和位置算符对检索项进行组配,编制成检索表达式。常用的布尔逻辑算符有“逻辑与”(and)、“逻辑或”(or)和“逻辑非”(not);常用的位置算符有“near”、“with”、“field”等算符。

3.2 选择检索工具

(1) 科研选题和科研进行中,由于要检索比较专业的、学术性较强的文献信息,如正式发表的期刊论文、会议记录等,应选择网上数据库检索系统,以保证检索结果的全面性和权威性。

(2) 网上商务信息数据库的比重逐年增加,如中文的万方数据资源系统和美国的 Dialog 公司均提供许多很有特色的商务信息数据库,用户如要检索比较准确和系统的商务信息(包括政策与法规、市场、金融、商品等),也可以选择网上数据库信息检索系统,登录其网站,进入有关数据库进行有偿信息检索。

(3) 检索时效性较强的信息,如新闻报道、最新商务信息等,可以选择网上搜索引擎。搜索引擎具有信息传递速度快且免费检索的特点,但对于学术

性强,比较专深的课题,用搜索引擎检索效果不理想。

(4) 针对具体的检索课题,可根据实际情况选用不同类型的信息资源灵活地配合使用,取长补短,即以专业性或综合性数据库检索为主,适当辅以网页搜索的检索方法,以达到最佳检索效果。实际操作中,用户最常用的就是这种综合运用的方法。

参考文献

- 1 黄如花. 网上电子期刊的利用. 图书情报工作, 2001(12)
- 2 陈光祚, 夏立新. 我国网络图书现状分析与发展对策研究. 中国图书馆学报, 2002(2)
- 3 李家清. 开发利用网络信息资源的对策研究. 图书情报知识, 2001(1)
- 4 司莉. 因特网上的图书馆虚拟文库建设. 图书情报知识, 1999(1)
- 5 李毅萍. 网络报纸资源及其开发利用. 图书馆论坛, 2001(4)
- 6 王建仑. 网络科研信息资源的开发和利用. 图书情报工作, 2000(2)

王云娣 浙江师范大学图书馆信息咨询部主任, 副研究馆员。通讯地址: 浙江金华。邮编 321004。

(来稿时间: 2002-09-19)