

● 刘晓清

全文数据库和电子文献传递系统研究

摘 要 全文数据库本身的存贮制式问题及标引和著录问题急需解决。建立电子文献传递系统可降低全文数据库的建设、收集和引进成本,逐步实现信息资源共享。实现电子文献传递系统的方式有:自动方式、人工辅助方式、网络访问方式和网络订购方式。参考文献 8。

关键词 全文数据库 电子文献传递系统 资源共享 网络化建设

分类号 G258.94

ABSTRACT The author thinks that we should solve the problems of formats and indexing of full-text databases. If we establish electronic document delivery system, we can lower the costs of development, collecting and purchasing of full-text databases and gradually realise resources sharing. There are four models of electronic document delivery system: automatic model, human-assisted model, network access model and network subscription model. 8 refs.

KEY WORDS Full-text database. Electronic document delivery system. Resources sharing. Networking development.

CLASS NUMBER G258.94

信息基础设施(信息高速公路)的建设在世界范围产生了极其重大的影响。在此基础上提出的国家知识基础设施(National Knowledge Infrastructure, 简称NKI)受到美国、西欧、日本等国家的高度重视,这是一项面向知识经济时代的重大信息工程。NKI的核心是建立知识库系统层,在此基础上通过信息基础设施和知识服务体系为各种知识应用提供支持,信息资源的建设、利用和研究则是知识基础设施的基础。作为信息资源的重要组成部分——全文数据库和电子文献传递系统也得到了越来越多的应

用和重视。研究和建立全文数据库及电子文献传递系统对我国图书馆的现代化建设具有非常积极的意义。

1 全文数据库的类型及制式标准

全文数据库(Full Text Database)又称源数据库(Source Database),是指由原始情报源构成的数据库,包含原始文献全文及专为用电子方法传递而编制的情报资料。从内容上分为数值型数据库、文字

为“普列汉诺夫”不是文献论述的和研究的对象和问题,故不是文献主题,不能出现在个人名称主题字段。

文献主题标引工作在我国开始于80年代,机读目录的编制工作还处于起步阶段,而对CNMARC格式的正确理解和确定规则,掌握主题标引的技术与方法,是保证计算机编目工作标准化的重要前提。为了实现网络环境下文献资源的共知共享,有关细则、标准亟待完善。

参考文献

- 1 北京图书馆《中国机读目录格式使用手册》编委会 中国

机读目录格式使用手册 北京:华艺出版社,1995

- 2 卜书庆 论计算机检索环境下的文献主题规范工作 北京图书馆馆刊 1998(2)
- 3 朱芊 也谈北图数据主题标引问题 北京图书馆馆刊 1998(2)
- 4 刘湘生 主题法的理论与标引 北京:书目文献出版社,1985

韩 敏 武汉大学图书馆编目部工作。通讯地址:武汉市。邮编 430072。

(来稿时间:1999-04-06。编发者:翟凤岐)

—数值型数据库、全文数据库三类;从载体上,分为磁带(磁盘)型、光盘型、网络型等。全文数据库的内容包罗万象,有财政、人口统计、辞典、手册、百科全书、期刊论文、法院判决、法律法规数据库等,如:大英百科全书、中国学术期刊(光盘版)全文库、中国专利说明书(光盘)等。当前,全文数据库的主要来源有二:一是现有馆藏的数字化。这是许多数字化图书馆工程的一项重要内容。大量的馆藏图书、期刊、图谱、古籍等通过光电扫描的方式转化为机读数据。这是一项庞大而持续时间又长的工程,需要大量的资金投入,如美国国会图书馆的馆藏数字化工程,上海图书馆的善本书数字化工程等。二是当前的计算机环境和网络环境下直接生产、转化并形成全文数据库,如中国学术期刊(光盘版)全文数据库等。

全文数据库的建立使我们获取一次文献的方式从传统载体转向机读载体,可以充分利用现代化的网络条件,及时、全面、准确地得到我们想要获得的资料信息。

随着全文数据库的发展,有两个问题急需解决,一是全文数据库本身的存贮制式问题,二是全文数据库的标引和著录问题。同时,由于目前全文数据库的出版形式多种多样,可以不受传统载体形态的限制,更给全文数据库的管理和利用带来了新的课题。当前,全文数据库的出版形式主要有:集成式、独立式和网络式。集成式,就是将多种期刊(或图书)集成在一个全文数据库系统出版发行和检索利用,如中国学术期刊(光盘版)、英国的ASTP(*Applied Science & Technology Plus*)都集成有几百到几千种期刊的全文数据,这种方式便于管理和利用,效率高,也是全文数据库的主要类型,面向图书情报机构、科研单位。独立式即每种期刊或图书单独发行全文光盘数据库,如《大英百科全书》多媒体版、《计算机世界》年度全文光盘、《北京大学学报》全文光盘等,它的出版方式灵活,更新方便,面向最终用户。网络式即直接在因特网上出版、发行或在印刷版出版的同时在网络上发行电子版的全文数据库,如网络小说、网络新闻、科技期刊电子版等。所有这些全文数据库的出版类型和存在方式都涉及到一个共同问题,即存储的数据格式——制式问题。分析当前的数据格式类型主要有:SGML——Standard Generalized Markup Language(标准通用置标语言)、HTML——Hypertext

Markup Language(超文本链接标示语言)、PDF——便携文件格式(美国Adobe公司提出的可移植文档格式文件)、图像格式(包括MAGE文档、TIFF文档、位图文档BMP、JPG文档、PCX/DCX文档、XIF文档等)、文本格式及各种专用格式(各出版公司自己定义的非开放的存储格式)。由于存储方式的多样化,就不可能以通用的全文管理系统(检索软件)对各种全文数据库进行管理和利用,因此,目前每一种全文数据库都有自己专用的全文管理系统,造成了新的影响数据交换和资源共享的障碍,使得最终用户必需掌握每种全文管理系统才能进行检索和利用。这就需要有一种大家都能遵循的全文数据库的存储标准和规范。

2 我国引进和自建的主要的全文数据库和中文全文数据库

2.1 引进的全文数据库

(1) *Academic Search Full Text Elite*(学术研究全文精英),1990始。该库收录了3200余种院校出版杂志(覆盖社会学、人文学、教育学等)的索引及文摘,其中1000余种杂志提供了全文,1700种杂志含同人评论。

(2) *IEEE/IEE*全文数据库检索系统:包括119种期刊、1815种会议录和部分标准,20多万篇文献覆盖了美国电气和电子工程师学会和英国电气工程师学会主要会议录和期刊。该系统收录文献时间范围从1988年到现在,每季度更新光盘。UMI公司出版。

(3) *Applied Science & Technology Plus*全文数据库:包括126种期刊全文,400多种期刊的摘要。专业涉及应用科学各领域。UMI公司出版。

(4) *UnCover*数据产品:这是美国CARL公司的主要产品,包括有期刊数据库和一系列有关信息检索和管理的集成化产品和服务。*UnCover*数据库目前收录期刊已超过18000种,其中约有51%属于科学、技术、医学和农林,40%属于社会科学、政法、商业,9%为艺术和人文科学。拥有期刊文章索引(或文摘)880万篇,并且每天以5000篇的速度不断扩充。

(5) *ProQuest Direct*(UMI公司)。*PQD*系统使用户能从网上多个数据库中连续检索,并随时获

取文章的全文。用户可以从系统提供的6种全文形式中选择最适宜的一种进行打印、下载或转发电子邮件。PQD 收集1987年以来的数千种期刊的全文和图像。24~48小时更新一次,提供最新的报纸、期刊全文。

(6) *ProQuest Research Library* (学术期刊图书馆), 1986始(UMI公司)。为“期刊文摘研究版”(PAR II)的升级版本,专为大学图书馆和研究图书馆设计的综合性学术期刊数据库。收录2308种综合性期刊和综合性报纸,其中有1488种刊是权威的Bill Katz的*Magazines for Libraries*里面所有的刊,1472种全文刊。所涵盖学科有:商业与经济、教育、历史、传播学、法律、军事、文化、科学、医学、艺术、心理学、宗教与神学、社会学及妇女感兴趣的题目。分为核心收藏和15个学科模块。核心收藏(core collection)包括上述学科领域的700多种核心期刊和重要学术期刊,15个学科模块(module)则是按学科对核心收藏的补充。目前可从网上检索到1971年来的文摘和1986年来的全文,每日更新。

(7) *Business Source Elite* (商业资料精萃)。该库收录31500余篇综合性文摘和索引,包括*The Wall Street*期刊和可以检索的840多种全文商业期刊。涉及的主题范围有国际商务、经济学、经济管理、金融、会计、劳动人事、银行等。EBSCO公司出版。

(8) *Facts on File* (事实数据库), 1980以后。该库收录美国和世界新闻的基本事件。

(9) *MAS Full Text Elite* (MAS全文精萃)。该库收录400多种一般性杂志和研究性杂志,还有140余种全文以及上千种的书评。最早期文摘始于1984年,最早的全文是从1990年开始。

(10) *World Magazine Bank* (世界杂志数据库): 收录澳大利亚、新西兰、北美和欧洲等国家400多种杂志的文摘和索引,还有155种全文杂志。提供的都是英语出版物。

(11) *First Search Base Package* 中的全文数据库。包括:*OCLC ArticleFirst*——12500多种期刊的文章索引,获取全文需另付费;*OCLC FastDoc*——带联机全文的文章索引,获取全文需另付费;*World Almanac*——世界年鉴;*H. W. Wilson Select*——H. W. Wilson公司的全文库,包括Wilson公司的普通科学文摘、人文学科文摘、读者指南文

摘和Wilson商业文摘库的资料,且检索全文无需另付费。

(12) *UK Vendor* 产品样本库: 包括英国及在英国设有代理机构的其他国家土木建筑工程的主要厂商(约4000多家)及其产品的全文信息。

(13) *CAPSP Expert* 电子元器件库: 提供了世界上最新的560万个以上的半导体、电子元器件、电子接插件。

(14) *SAE MOVE* 光盘数据库: 是美国机动车工程师协会出版的光盘数据库产品,包括有论文全文、标准、技术规范等。

2.2 中文全文数据库

(1) 中国学术期刊(光盘版)全文数据库: 收集3500多种1996年以来的中文核心期刊全文数据库,学科范围包括理工、农林医、社会科学、教育等领域。

(2) 中国人民大学复印报刊资料全文数据库: 包括1995年至现在的“人大复印资料”的全部信息。

(3) 中国专利说明书全文光盘: 包括中国专利局公告的专利说明书全文。

(4) 《中国大百科全书》多媒体光盘。

(5) 《人民日报》《经济日报》《参考消息》《解放军报》等近200种报纸全文光盘。

(6) 《四库全书》原文电子版: 包括经、史、子、集四部,计153张光盘。

(7) 各种电子图书。

3 建立电子文献传递系统的意义

随着网络的使用和普及,人们对信息的获取越来越多地依赖于网络环境。目前,二次文献都可以通过网络得以检索和获得。但全文数据库作为一种源数据,并不适合全部上网,在网上阅读一篇宏篇巨著并把它打印出来是非常困难的,以传统的手段,人们可以通过馆际互借、代办复制等方法解决,但由于其时效性、方便性较差,不能令人满意。因此,在因特网给信息用户带来便利的同时,用户更希望在信息获得方式上得到“一站式”服务,即在检索到文献信息线索以后,立即(或尽快)获得该信息的原始文献(全文数据库),免除再通过另外的方式和渠道去找寻原始文献的麻烦。中国教育科研计算机网络(CERNET)和中国高等教育文献保障体系(CALIS)两个项目的实施,建立了我国教育信息基

基础设施和文献信息资源的保障体系。在项目建设中,数据库的建设尤其是全文数据库的引进和有我国特色的全文数据库的建设得到了高度重视。为了使全文数据库的建设、收集和引进降低成本,逐步实现信息资源的共享,必须建立相应的资源共享的应用环境,即电子文献传递系统,充分发挥网络基础设施在信息传输方面的优势。建立一个遍布全国的全文数据库分布式保障体系,使一次文献资源的共享得到技术上的充分保障。

4 电子文献传递系统的实现方式

图书馆的网络化和自动化建设使图书馆建立了较为成熟的图书馆内部资源网络和虚拟图书馆网络,在其信息资源管理平台融合了因特网、万维网、HTML 超文本信息链等图文声像并茂的多媒体开放文档体系结构,交互式对象的中西文全文检索等各项当今最新的技术,将多种平台的服务器组成一个巨大、开放的虚拟图书馆。信息资源管理平台方便、适用的全文信息检索系统和电子文献传递系统,集数据库管理技术、全文信息管理技术、中文信息处理技术及电子文献传递技术于一体,在因特网上运行,满足信息服务、信息传递和信息发布的需求。

4.1 自动方式

信息用户在网络上通过二次全文数据库或搜索引擎获得检索结果后,可以直接选择“订阅全文”或类似功能向全文数据库管理系统发出订购信息,电子文献传递系统在对用户进行资格认证、结果检查,电子付款后,将存储在全文数据库(在光盘或磁盘阵列)中的命中文献以电子邮件方式发送给用户,用户收到电子邮件后通过解压即可阅读和处理全文。

4.2 人工辅助方式

用户在获得文献线索后,通过联合目录获知该全文数据库的收藏馆,通过电话、传真或电子邮件通知收藏馆,收藏馆根据需求将全文数据以传真或电子邮件方式发送给用户。

4.3 网络访问方式

用户通过万维网浏览器,获得文献线索,通过授权登记后直接在网络上阅读或下载全文数据。

4.4 网络订购方式

用户根据自己的需要,在WEB站点上公布的专题信息资料库中选择自己感兴趣的专题信息,将

自己的电子邮件加入到订购的邮件列表中,一旦有新的信息加入,WEB站点便通过邮件列表将信息发送给用户。

建立电子文献传递系统,必须有相应的支撑系统相配合。除了网络基础设施外,应建立一套全文数据库的存储规范,编制全国的全文数据库联合目录,开发一套电子文献传递管理系统的软件,同时应有电子商务系统的支持和配合。

互联网为人们以便捷方式查询信息提供了机会,通过超文本系统和HTML编辑,可将各种全文数据库信息集中在一个或多个联合目录上供给用户,尽管这些资源分布在世界各地的主机上,但互联网络可以不受地理和时间上的限制。这正是分布式全文数据库和电子文献传递系统的优势所在。要实现全文数据库的资源共享,应着重研究和解决全文数据库(包括电子图书)的存储制式规范和统一问题,开发通用的电子文献传递系统的规范和标准,包括传递协议、检索软件和管理系统问题以及寻找和采用电子货币结算的方式和方法,即取得电子商务系统的支持的问题。

参考文献

- 1 吴慰慈 网络时代文献资源共享的新模式: 电子文献传递 中国信息导报, 1996(5)
- 2 叶继元 电子期刊收集策略探微 中国图书馆学报, 1998(5): 51~ 54
- 3 朱强 英国高等学校的信息资源共享 大学图书馆学报, 1998(6)
- 4 董小英 信息服务: 网络环境下的热点 计算机世界, 1998(7)
- 5 董慧 信息网络全文数据库设计的思考 情报学报, 1998(1)
- 6 何小清 美国图书馆网络对信息资源的建设与组织给我们的启示 中国图书馆学报, 1998(6): 67 ~ 72
- 7 真溱, 胡从玉, 胡均平, 王新 面向网络环境的图书馆模式构想 中国图书馆学报, 1998(6): 47~ 50, 82
- 8 <http://uncweb.carl.org/uncover>

刘晓清 浙江工业大学图书馆副研究馆员。通讯地址: 杭州浙江工业大学图书馆。邮编 310014。
(来稿时间: 1999-04-22。编发者: 徐苇)