

●唐健雄¹, 李世玲²

(1. 河北师范大学 文学院, 河北 石家庄 050091; 2. 河北师范大学 图书馆, 河北 石家庄 050091)

集成信息检索系统的特点

[关键词] 集成; 信息检索系统; Web of Knowledge

[摘要] 分析了信息检索的发展历程, 以 ISI 提供的 Web of Knowledge 为例, 讨论集成信息检索系统的特点。

[中图分类号] G252.7

[文献标识码] A

[文章编号] 1005-8214(2005)01-0099-02

1 引言

信息化社会, 人们涉猎的信息越来越广泛, 需要从各种各样的信息资源中获取有关的信息。图书馆和信息服务机构面对这样的形势, 要满足用户的多种信息需求, 有效利用各种信息资源, 有必要建立集成信息检索系统。这个集成信息检索系统就是把所有支持用户检索信息服务工作的多个系统集中在一起, 共同实施信息服务工作的方法。例如, ISI (Institute for Scientific Information, 美国科技情报所) 提供的 Web of Knowledge 就是为教育和科研人员检索高质量学术信息的综合平台。^[1] 此系统具有丰富的、高水平的信息内容, 并且有检索各种有效信息内容的高级检索工具, 引导用户快速获取有关的信息内容。本文以 ISI 提供的 Web of Knowledge 为例, 讨论集成信息检索系统的特

与, 加大建设数字图书馆的力度。

●国外数字图书馆的项目已经发展到了综合应用的阶段, 其起点高, 投入大, 社会基础较好, 相比而言, 我国各方面条件都远远不足, 因此在进行数字图书馆建设时, 按照规模大小和功能强弱可以分为几个不同的层次: 单机环境下的数字图书馆 (个人或部门层次)、局域环境下的中小型数字图书馆 (机构层次)、广域环境下的专业数字图书馆 (系统或地区层次) 和广域环境下的综合数字图书馆 (国家层次)。^[11]

[参考文献]

- [1] 雄虹, 艾润莲, 李丹. 我国数字图书馆建设模式的研究 [J]. 现代图书情报, 2002, (4): 7-10.
- [2] 刘炜. 数字图书馆概要 [EB/OL]. <http://www.libnet.sh.cn/istis/dlib/report/dlibintro.html>, 2004-11-12.
- [3] 上海图书馆数字化课题组. 美国数字图书馆首倡计划 [EB/OL]. <http://www.libnet.sh.cn/istis/dlib/report/DLI.html>, 2004-11-12.
- [4] 阮延生. 数字图书馆知识产权问题研究 [J]. 中国图书馆学报, 2002, (1): 54-56.

点。

2 信息检索的发展历程

随着计算机和信息技术发展潮流的影响, 人类检索信息经历了从手工检索到机器检索, 又从机器检索向机器引导的方向发展。Stern 把信息检索的发展历程分为印刷媒体检索、第一代联机 (磁盘、光盘) 检索、第二代 (网络) 检索。^[2]

2.1 印刷媒体检索

印刷媒体检索指根据卡片式目录和书本式目录、缩微型目录、文摘、索引等方式的检索。作为有效的检索手段, 研究制定出了统一用语和各种分类号。

2.2 第一代联机 (磁盘、光盘) 检索

20 世纪 70 年代, 利用先进的卫星通信技术, Dialog、Orbit、Medline 等联机检索系统开始跨地区为世界各地的远程终端用户提供联机检索服务。联机检索计算机除可以检索出所有的记录项目之外, 也可以使用布尔运算符检索, 从而实现更有效的检索。但是使用主要的公用线路, 传统的联机检索需要获取检索指令和花费较高的检索费用, 同时还要具备利用检索系统相关的知识和技术。

20 世纪 80 年代, 随着激光存储技术的广泛应用, 促使 CD-ROM 问世, 继之磁盘不断被光盘代替, 信息成本大大降低, 信息检索也更加接近终端用户。Windows 等的

- [5] 吴苻. 数字图书馆建设中的知识产权问题 [J]. 图书馆工作与研究, 2002, (3): 29-32.
- [6] 赵继海. DRM 技术的发展及其对数字图书馆的影响 [J]. 大学图书馆学报, 2002, (1): 14-16.
- [7] 陈海英. 数字图书馆中的数据压缩及信息共享问题 [J]. 大学图书馆学报, 2002, (2): 19-21.
- [8] 付昕, 耿敏. 数字图书馆中的视频信息检索 [J]. 大学图书馆学报, 2002, (2): 22-24.
- [9] 任瑞娟. XML 对数字图书馆的影响 [J]. 大学图书馆学报, 2002, (6): 13-16.
- [10] 刘炜, 刘年娣. 国内数字图书馆研究述评 [EB/OL]. <http://www.libnet.sh.cn/istis/dlib/report/SZ1.html>, 2004-11-12.
- [11] 孟连生. 关于发展我国数字图书馆事业的几点思考 [J]. 图书情报知识, 2002, (1): 9-12.

[作者简介] 束漫 (1971-), 女, 广州大学图书馆副研究馆员, 中国科学院文献情报中心博士研究生, 已发表论文 10 余篇。

[收稿日期] 2003-08-30 [责任编辑] 师宏睿

出现,使得终端用户能够很容易地在个人电脑上进行信息检索。但是,由于各链接的接口不同,造成每个数据库操作方法各异。而联机检索主要是根据主框架的单一检索系统来实现多个数据库的交叉检索,这就需要以一定的形式,把不同种类数据库的索引和主题词表合并在一起才能实现。对异种数据库的交叉检索,现在正广泛研讨Z39.50议定书的检索。与Z39.50对应的图书馆OPAC系统和各种文摘、索引数据库为数很多,即使是Z39.50也仍然存在着检索项目受限制等方面的制约。^[3]

2.3 第二代(网络)检索

20世纪90年代,计算机网络技术迅猛发展和元数据不断完善,促使信息机构开发出了网络检索平台,提供网络版数据库。如ISI Web of Science、ISI Proceedings、ISI Current Contents Connect、ISI Journal Citation Reports、ISI Chemistry、DII (Derwent Innovations Index)、BIOSIS Previews和INSPEC等等。与第一代联机检索主要依靠信息处理技术的发展相比,第二代信息检索要依靠信息技术和元数据双方面的支撑。在技术方面由于因特网和Web字符集的进展,由一个信息可以找出相关的信息来。第二代(网络)检索的显著标志是依靠Web技术,ISI的Citation Index作为ISI Web of Science[®]提供,它是根据概念之间的连接进行自然语言检索的模式,即用包含主题术语或关键词的概念检索。另外,还有SDI、简便的用户检索接口、最合适的排序、全文检索、人工智能、目标指向接口、超链接技术等。为实现检索学术信息的完整性和高质量,ISI公布了集成信息检索平台——ISI Web of KnowledgeSM。

3 集成信息检索系统的特点

3.1 具有丰富的信息内容

集成信息检索系统包含的信息内容的质与量是非常重要的,为满足用户多种多样的信息需求,需要重视信息内容的广泛性,详尽的信息目录及高品位的学术价值。如ISI的Web of Knowledge包括以下种类的信息源。

3.1.1 ISI公司、Derwent公司制作提供的信息

第一种信息源是ISI或其姐妹公司Derwent制作的数据库,如ISI Web of Science、ISI Proceedings、ISI Current Contents Connect、ISI Journal Citation Reports、ISI Chemistry、DII等,具体有学术杂志论文、会议记录、化合物信息、专利信息、Web目录等多种类型。从收录年代看,有1945年以后的杂志论文的书目信息,化合物信息;最古老的有1840年以后的数据。收集的最新的学术信息多数是Web上流通的,Web文献也是ISI检索信息源的采集对象。

3.1.2 ISI主要提供的集成数据库信息

第二种信息源是ISI平台提供的其他机构制作的数据库,如BIOSIS Previews^R(生物科学),CAB Abstracts^R(农业科学),INSPEC(物理、工程科学)等在学术界颇具影响的数据库。因为Web of Knowledge平台吸收了这些数据库的信息,从广泛集成ISI独立的数据库,可以更加深入地收集某个专业领域的信息。

3.1.3 链接外部数据库信息

第三种信息源是能够集成利用外部机构提供的信息,如Web of Knowledge平台与电子期刊全文链接,可以浏览电子期刊全文内容;与馆藏目录OPAC链接,可知某种文献是否在该图书馆收藏;与其他数据库链接,可知该数据

库信息以及相关信息。对此,在后面谈到的异种数据库的横断检索时要详细地论述。

3.1.4 分析、评价信息的工具

第四种信息源是ISI的优势或者特长,即具有累计引用信息和分析、评价信息的功能。使用Impact Factor能够广泛了解引用信息方面的情况,Journal Citation Reports^R是分析、评价杂志的工具,应用ISI Essential Science IndicatorsSM能够根据各学术领域的引用动向探讨某个领域的研究动向。

Web of Knowledge包括的信息内容如上所述,但因Web of Knowledge具有后面谈到的与相关信息的链接功能,实际上在Web of Knowledge平台上检索到的信息大大超出了上述范围。

3.2 提供高水平的学术信息

研究人员在搜索信息时,不仅希望能够找到信息,而且能从丰富多样的信息中迅速找出所需要的信息。从这一点考虑,集成信息检索系统不仅要有广泛的信息内容,而且这些内容是经过组织、筛选后选择出来的对用户需求价值高的信息,这样才能保障提供有效地信息检索服务。如ISI对提供的各种信息资源制定了严格的选择标准,根据标准系统编集信息资源,对学术研究人员只提供有参考价值的信息。^[4]现在,学术信息在Web上公开的很多,在各研究机构的位置除介绍研究活动外,还准备发行手册,这对探讨最新的研究动向发挥了很大的作用。与此同时,由于因特网和Web技术的普及,学术信息的流通信道也有了很大的改进。

在这样的背景下,ISI为真正能够提供最新的学术信息,把Web上公布的各研究单位的网址,如ISI Current Contents Connect(学术网址数据库),或者文献资料都被列为采集的对象,采集的Web信息也和学术杂志一样,具有严格的选择标准。所以,在Web of Knowledge平台上能够迅速找到学术价值高的Web信息。^[5]这样,多种多样杂乱无章的信息经过筛选再提供,大大缩短了研究人员从丰富的信息源中找出有价值信息的时间。

3.3 异种数据库的交叉检索

异种数据库的交叉检索是集成信息检索系统应具备的必要功能。数据库的数量很多,并且各具特点,即使在同一领域也存在着多个数据库。进行边缘领域的学术研究时,需要从边缘领域数据库及多个专业领域的数据库全面地进行信息收集。如前所述,从联机检索的初期阶段就开始实行多个数据库的交叉检索,但那时的交叉检索是把数据库的构造统一起来实现的,从某种意义上来说,是在“同种数据库(homogeneous databases)之间的交叉检索”。但是,随着信息技术的发展,在不同构造的数据库之间也能够进行“异种数据库(heterogeneous databases)的交叉检索”,这种异种数据库交叉检索将逐渐用Federated Databases语言表示。

ISI的Web of Knowledge具有CrossSearch功能,可以收集多个数据库进行检索。CrossSearch检索的信息内容,不仅有ISI公司生产的信息资源,还有ISI主要提供的其他机构生产的信息资源、外部机构提供的信息资源、Web上的信息资源等。CrossSearch使用的检索引擎,因为无意识记录的构造,像书目记录构造化的数据、(下转第106页)

文科图书引进专款, 亚基会通过大连理工大学、青岛海洋大学、上海同济大学等转运站提供的国外捐赠图书, 对于我们西部地方综合大学图书馆无疑于雪中送炭。

(5) 有偿服务增加经费来源。西部地方综合大学图书馆应该充分利用本馆的文献资源和学校的智力资源, 通过对馆藏文献的开发, 开展有偿服务或创收活动, 从所得收入中拨出一部分资金来补充文献购置费, 支持文献资源建设。

加强文献资源建设, 首先要抓采访队伍的建设。要选拔综合素质较高的人员充实采访队伍。在管理机制上, 应重视采访工作人员的工作职责与权限的规定, 严格建立采访验收制度。有条件的馆, 可学习一些重点大学提出的学科馆员采访制度, 提高购入文献的学术性和针对性。对采访工作中的一些原则, 如各类文献的经费分配, 各级文献的收藏比例, 文献收藏范围, 入藏文献的分配、布局, 文献资源的共建共享等, 都应有明文规定, 以便工作人员操作和自我约束。要促使工作人员从以往的凭书目订单的事务性采访工作活动中解脱出来, 更多地转向广泛收集各种信息, 利用采访网络向专家学者咨询, 加强与学科带头

(上接第 100 页) pdf 形式、html 形式、教科书形式等等, 所有记录符号的信息都能够进行全文检索。检索方式除过去指定的字段检索格式之外, 也能够根据自然语言进行概念检索, 充分显示出电脑的高智能, 信息检索的简单化。

ISI 为进一步提高异种数据库的检索功能, 与 WebFeat 公司合作, 开始使用 Feat Prism 提供异种数据库的检索, Feat Prism 能提供超过 1200 个各种数据库的交叉检索。

3. 4 与相关信息链接

集成信息检索环境中具有导航相关信息的功能, 根据检索结果或者已知的信息追寻导航 (链接) 相关的信息, 从而获得更多的信息。与相关信息的链接, Web of Knowledge 提供如下三种链接方式。

3. 4. 1 产品内链接

像第二代联机 (网络) 检索那样, Web of Knowledge 提供的引文导航是根据自然语言检索, 通过处理概念之间的关系找出相关信息的有效方法。如 Web of Knowledge 准备了“某文献引用什么?”“某文献如何被引用?”“有使用和某文献相同的引用文献的文献吗?”三种引文链接, 使用这些链接可以在短时间内找出所有与研究题目有关的文献。

3. 4. 2 产品之间的链接

产品之间的链接是指实现 ISI 生产的信息源或者主要提供的信息源相互之间的链接。这种链接的优点如下:

能够从多个角度探讨对一个信息的看法。例如, 发现某个有意义的信息时, 刊载该信息的杂志在学术领域具有怎样的特征? 该杂志还刊载哪些论文? Web 信息中有该信息的题目记录吗? 从而可以改变信息检索点等。

ISI 主要提供的信息源可以检索 Web of Science 的引文信息。例如, 根据收录在 INSPEC 的文献信息, 能够追寻该文献的各种引文, 从而提高信息的附加值。

3. 4. 3 与外部数据的链接

例如, ISI 数据库的书目信息中有电子杂志全文的链接, 通过此链接可以使用文献数据库阅读发现的有意义的文献。根据与 OPAC 链接功能, 可以马上知道该论文刊载的杂志是

人、专业教师之间的交流和沟通, 提高采访决策的科学性、针对性、实用性。

[参考文献]

- [1] 刘桔. 中心城市地方大学图书馆藏书结构探析. 佛山大学学报, 1998, 16 (1): 91-96.
- [2] 吴志荣. 实体图书馆与虚拟图书馆相结合的新型图书馆. 现代情报, 2003, (7): 116-117.
- [3] 中华人民共和国教育部. 普通高等学校本科教学水平评估指标 [Z]. 2004-08-12.

[作者简介] 张向东 (1946-), 男, 宁夏大学图书馆研究馆员; 孙方礼 (1949-), 男, 宁夏大学图书馆副研究馆员; 崔润新 (1974-), 男, 宁夏大学图书馆馆员; 侯丽君 (1958-), 女, 宁夏大学图书馆副研究馆员、采编部主任。

[收稿日期] 2004-05-24 [责任编辑] 周天昊

否在某图书馆收藏。从 BIOSIS Previews 数据库可以获得生物学方面的信息, 从 INSPEC 数据库可以了解物理、工程学方面的信息, 从 DII 获得某项研究的专利信息等。

3. 5 其他功能

Web of Knowledge 把支援科研工作作为自己的神圣职责, 不仅检索和导航信息, 而且使用有效收集信息的高级工具, 扩大 Web of Knowledge 的功能。如从各种信息源中发现书目信息时使用 EndNote 和 Procite, Reference Manager 文献管理软件可以很容易地下载。

[参考文献]

- [1] 林佳, 杨毅. 文摘索引型数据库检索系统的现状与发展趋势 [J]. 图书情报工作, 2003, (10): 68-73.
- [2] Stern, David. New search and navigation techniques in the digital library [J]. Science & Technology Libraries. 1999, 17 (3/4): 61-80.
- [3] Ferrer, Robert. University of Illinois the federation of digital libraries: interoperability among heterogeneous information systems [J]. Science & Technology Libraries. 1999, 17 (3/4): 81-119.
- [4] Cronin. The ISI Web of Knowledge platform: current and future direction [EB/OL]. <http://www.isinet.com/essays/issplatform/8105318,2001-03-24/2001-08-17>.
- [5] Eugene Garfield. Current web contents: developing web site selection criteria [EB/OL]. <http://www.isinet.com/essays/selection of material for coverage/23.html,2001-05-10,2001-09-20>.

[作者简介] 唐健雄 (1954-), 男, 河北师范大学文学院教授、硕士生导师, 发表论文 20 余篇; 李世玲 (1954-), 女, 河北师范大学图书馆副研究馆员, 发表论文 20 余篇。

[收稿日期] 2004-02-23 [责任编辑] 周天昊