

学科信息门户信息组织的评价

黄如花

(武汉大学 信息管理学院, 湖北 武汉 430072)

[作者简介] 黄如花(1968-), 女, 湖北新洲人, 武汉大学信息管理学院图书馆学系副教授, 博士, 主要从事信息检索、网络信息组织研究。

[摘 要] 学科信息门户信息组织有其自身的评价指标体系, 即学科资源的覆盖率、资源的质量和对资源的描述、更新与维护、互操作性、相关技术的运用、词表和安全性能等, 在实际运用中, 尚应做必要的调整和加强。

[关键词] 学科信息门户; 信息组织; 评价

[中图分类号] G254 [文献标识码] A [文章编号] 1671-8828(2003)05-0653-05

互联网上最流行的信息组织工具——搜索引擎确实在一定程度上解决了网络信息的序化问题。但由于没有采用受控语言, 它们的检准率低(对学术信息的检索尤甚), 又难以检索出隐性信息(如 PDF 文件、保存于后台数据库中的文档等), 索及的范围也相当有限。根据美国总统信息技术咨询委员会 2001 年 2 月的报告所引用的数据, 目前还没有一个搜索引擎的索引量能够超过万维网所有公共可索引信息的 16%。另据统计, 所有搜索引擎覆盖的学术信息不足网络学术信息的 1/3。因此, 对于查找某学科(专题)的学术信息而言, 搜索引擎显得力不从心。在这种背景下, 图书情报界作为印刷资源的主要组织者, 也加入了网络信息组织的行列, 他们纷纷将图书馆传统的信息采集、标引和组织优势扩展到 Web 空间, 试图提高网上资源的序化程度, 以弥补搜索引擎的不足和改善目前检索低效的状况, 开发出了“基于学科的信息门户”(Subject Based Information Gateways, 即 SBIGs), 本文简称为“学科信息门户”。

SBIGs 曾经是“提供可检索和可浏览的因特网资源目录的联机服务系统, 集中于某一相关的学术领域, 提供对经图书馆工作人员遴选和按学科组织的因特网资源的利用”。在开放式数字信息服务环境下, “学科信息门户致力于将特定学科领域的信息资源、工具与服务集成到一个整体中, 为用户提供一个方便的信息检索和服务入口”^[1](第 88 页)。

(一) 学科信息门户信息组织的特点

学科信息门户按照某学科(专题)用户的要求对网络中相关的信息资源进行了更有针对性、更深入的揭示, 在给用户“指路”的同时提供更专门、更专深的信息检索服务, 有助于专业用户在本领域的“信息超市”(information supermarket)中选择高质量的资源和获得“一站式检索”(one-site search), 从而保证用户获得“所得即所要”的信息。对图书馆来说, 学科信息门户拓宽了本馆的“馆藏”, 而对整个网络而言, 其信息的序化程度得以提高。

学科信息门户的信息组织需要图书情报学、计算机科学和相关学科专家的共同参与, 才能保证其质量。下表中的 SOSIG 便是图书馆学专家、信息科学专家、计算机科学专家对图书馆应用、实践与数据库技术有机结合的结果, 旨在为社会科学研究人员、大学教师和图书馆员提供世界范围的因特网资源的快

捷而高质量的检索,一投入使用便大受欢迎。它提供的每一种资源都经过图书馆员和学科专家的选择与描述,收录的资源有:电子期刊、数字化图书、报告、论文、书目、教育软件、电子新闻稿、邮寄单和重要的社会科学机构的主页。

正是由于学科信息门户具有这些优于搜索引擎的特点,新的学科信息门户近几年在国外不断涌现。2001 年底正式启动的中国国家科学数字图书馆(The Chinese Science Digital Library, CSDL)已建立起图书情报学等多个学科信息门户。国内外重要的学科信息门户见表 1。

表 1 国内外重要的学科信息门户

国家	门户名称	学科	开发者	URL
美国	WWW Virtual Library	综合	HTML 的创立者 Tim Berners-Lee 和志愿者	http://vlib.org , 在英、阿根廷、瑞士有镜像站点
美国	Argus Clearinghouse for Subject-oriented Internet Resources	综合	Argus Association	http://www.clearinghouse.net
美国	INFOMINE	综合	加州大学、底特律大学等	http://www.clearinghouse.net
美国	Isaac Network	跨门户	威斯康辛大学麦迪生分校, NSF 资助	http://scout.cs.wisc.edu/research/isaac/index.html
英国	CrossROADS	跨门户	经济与社会研究协会	http://www.ukoln.ac.uk/metadata/roads/crossroads/
英国	Social Science Information Gateway (SOSIG)	社会科学	经济与社会研究协会、欧盟	http://www.sosig.ac.uk
美国	Librarians' Index to the Internet (LII)	图书情报	加州大学图书馆和加州其他馆	http://lii.org/
英国	Bulletin Board for Libraries, Libraries of Networked Knowledge (BUBL LINK)	图书情报	合作信息系统委员会 (JISC)	http://bubl.ac.uk/link/
中国	CSDL 学科信息门户系列	图书情报, 数学、物理、化学、生命科学、资源与环境	科学院	http://www.cSDL.org/index.jsp
美国	Education Resources Information Center (ERIC)	教育	教育部	http://ericir.syr.edu
澳大利亚	Education Network of Australia (EdNA)	教育	教育部和高等教育机构	http://www.edna.edu.au
英国	Edinburgh Engineering Virtual Library, (EEVL)	工程	爱丁堡大学与其它 6 所大学	http://eevl.icbl.hw.ac.uk
澳大利亚	Australian Virtual Engineering Library (AVEL)	工程	工程研究所和皇后岛大学等	http://www.avel.edu.au
澳大利亚	Agrigate	农业	墨尔本大学等	http://www.agrigate.edu.au
澳大利亚	MetaChem	化学	新南威尔士大学等	http://metachem.ch.adfa.edu.au/
美国	The Art, Design, Architecture & Media Information Gateway (ADAM)	艺术	ADAM 联盟	http://adam.ac.uk/
美国	Virtual Business Information Center (VBIC)	商业与经济	马里兰大学商学院、信息研究学院和图书馆	http://vbic.umd.edu
英国	BIOME	生命与健康科学	牛津大学等	http://biome-ac-uk

(二)学科信息门户信息组织的评价指标

为了优化现有的学科信息门户,并为新的学科信息门户建设提供参考,需要对学科信息门户的信息组织提出一套评价指标。

1. 学科资源的覆盖率。是否汇集了某学科(专题)领域重要的相关资料?门户中收录的资源类型是否多样?是否能方便用户对某一专题信息资源的“一站式检索”?一个好的学科信息门户所指引的有关资源应尽可能地涵盖本领域重要的资源。如,INFOMINE的记录达4万余条。

理想的学科信息门户应该从仅收录因特网资源发展到囊括馆藏实体资源(包括二次文献数据库、全文数据库、馆藏目录、联合目录等)的集成化系统,两种资源在同一界面实现无缝存取(seamless access),整合为易检易用的有机整体,即“hybrid libraries”(复合图书馆)。基于用户利益而管理的复合图书馆:用于管理与组织的人文科学复合图书馆模型(MALIBU)便是一个成功的范例。

2. 资源的质量和对资源的描述。指引的资源是否适合目标对象?对资源的描述是否准确?航标是否明确?指向的资源地址是否正确?是否链接资源的原始网址以便用户访问到最及时和最权威的信息?在可存取性和可用性方面,是否考虑到用户访问本门户网站采集的资源时所需的硬件、软件和连接方式等要求?当网络用户与镜像站点的“电子距离”比原始站点更近时,是否链接了镜像站点以便于用户进行更有效的存取?学科信息门户对有关网站的分类要深入,所指引的网址要尽量接近主题内容,尽量减少用户点击的次数。

对资源的描述是否采用国际通用的元数据、标记语言、分类法与词表?如澳大利亚的SBIG全部采用DC和澳大利亚政府查找服务(Australian Government Locator Service, AGLS)元数据元素以及自己独有的元数据元素,并支持因特网内容选择平台(Platform for Internet Content Selection, PICS)。INFOMINE的记录以HTML语言表示,主题标引使用LCSH,记录还可以转换成MARC格式的记录。

3. 更新与维护。由于网上站点的增加与更改频繁,学科信息门户建设中的更新与维护不容忽视,这也是保证学科信息门户质量的最重要的后续工作。学科信息门户要真正成为因特网信息的深层组织工具,必须改变目前因追求高质量而过于依靠人工参与的状况,充分利用网上自动漫游、自动跟踪、自动分类和自动标引技术,采用人机结合的工作方式,为用户提供更优质、高效的服务,还可以鼓励用户推荐资源或参与维护(如www.Virtual Library和Agrigate)。

更新是否及时?是否有专门的工作人员不断追加新的网络资源?是否及时剔除错链、死链?整个系统在结构上是否为一个活的系统?是否可根据用户需求和网络信息资源的变化及时对词表和分类体系进行调整?维护是否比较容易?软件是否可靠?

瑞典工程学电子图书馆由于人工编制索引,跟不上网络资源变化的速度而不得不停止使用,计划在以机器人编制索引的All-Engineering基础上开创一个新的门户。

4. 互操作性。学科信息门户中各信息源数据库与信息平台差异可能很大,为了在统一的界面使用来源各异的网络资源,学科信息门户必须具有异构计算机软硬件平台间良好的互操作性,具有跨门户检索的能力。

为了提高互操作性,美国国家自然科学基金会(NSF)资助的Isaac Network采用DC作为元数据,以Linux作为平台,以Lightweight Directory Access Protocol(LDAP)作为信息查询与交换软件,而以代号为RFC2651的通用索引构建协议(The Architecture of the Common Indexing Protocol, CIP)作为索引编制与互换协议。通过该门户,用户可以对SOSIG, BUBL LINK, EEVL, EdNA, MathGuide, GeoGuide, OMNI等近20个学科信息门户进行跨门户的检索。同样由NSF资助的“全国科学、数学、工程和技术教育数字图书馆”(SMETE)项目则将多个分布式学科信息门户作为整个数字信息资源的整合机制和服务渠道,允许用户通过该门户体系检索和调用各种不同的信息资源与服务。

5. 相关技术的运用。学科信息门户的运作仅靠图书馆的理念是不够的,要涉及到大量的实现技术,如虚拟现实(Virtual Reality, VR)技术、虚拟专用网(Virtual Private Networks, VPN)技术、虚拟局域网

(Virtual Local Area Network, VLAN)技术、虚拟数据库(Virtual Database, VDB)技术、通用对象请求代理体系结构(Common Object Request Broker Architecture, CORBA)技术等。例如,利用 VPN 可以解决学科信息门户信息共享的安全问题。信息推送(Push)技术通过信息代理机制,在用户初次使用时设定所需的信息后,能通过推送(Push)或网播(Net-casting)的方式把网上相关信息送到用户面前。这种基于 Push 技术的 Internet 信息检索技术既为用户搜索、浏览网上的相关信息提供了快捷入口,又为学科信息门户在广域网内的信息共享提供了技术支持。SOSIG 采用了自动标引、分布式编目、镜像等先进技术。

6. 词表。学科信息门户词表的设置与否对于本专业的深入检索影响很大。学科信息门户应建立符合该专业特点的主题词表和分类体系。例如,上表中的 ERIC 建立的叙词表是一个层次结构的术语体系,提供与某一术语更广或更窄的术语及其相关术语间的链接。SOSIG 建立了社会科学的叙词表以帮助用户限定检索,该词表以英国 Essex 大学开发的“人文科学与社会科学电子词表”(Humanities and Social Science Electronic Thesaurus, HASSET)为基础。Agrigate 采用 CAB 国际农业词汇(CAB International Agriculture Thesaurus)进行元数据创建、主题浏览和主题检索。

7. 安全性能。数字化方式存贮的信息,极易受到计算机病毒、黑客入侵等的干扰和破坏,资源的安全受到了许多挑战。学科信息门户的安全性问题应该受到重视,要注意对资源进行备份与保存。

评价学科信息门户时还应该考虑用户界面的友好性和服务功能的拓展。

(三)学科信息门户实例 LII 评价

“图书馆员因特网索引(Librarians' Index to the Internet, 简称为 LII)”是将图书馆组织信息资源的方法利用于学科信息门户的典型。它始于加州大学伯克利分校参考馆员 Carole. Lelta 1990 年制作的 Gopher 书签文档,1993 移到伯克利公共图书馆 Web 服务器,命名为 Berkeley Public Library Index,1997 年又移到伯克利数字图书馆 SunSITE 网站,更名为 Librarians' Index to the Internet。LII 是一个可供检索的、提要性的学科信息资源目录,收录 30,000 多个经过选择和评价的网站,基本收录了网上重要的图书情报类资源,指引的资源非常适合公共图书馆用户、图书馆员和图书情报领域的研究者。

LII 选择与评价资源的标准是:最适合公众和图书馆用户需求的资源,不收录没有信息内容的纯粹商业站点,而且,每一种资源都经过图书馆员对其价值进行评价。Carole. Lelta 于 2001 年 10 月退休后,由 Karen G. Schneider 全面负责,有 4 位兼职工作人员,另有加州范围内 100 多位图书馆员志愿者,工作人员都是受过训练的,也注意培训新的索引人员。这样,可以确保资源的质量。这里的资源被分为艺术、商业、计算机、文化、教育等约 40 个大类,大类下再分小类。每条记录都给出了详细的介绍,包括资源名称、URL、摘要、主题词、记录创始人与时间,以及记录修改人与时间。摘要由 LII 工作人员撰写,使读者在进入一个网站前便可了解其主要内容。LII 链接的基本是资源的原始地址。对工作流程、著录项目与规则以及注意事项均做了非常具体的规定,因而能够保证资源描述的质量与一致性。

LII 采用国际通用的《国会图书馆主题词表(LCSH)》进行标引,但又将 LCSH 中的部分主题改为更符合公众特点的主题名称(例如,在 LCSH 中用 electronic mail system,在 LII 中用 e-mail)。它遵守 Z39.50 与 MARC 标准,使得它具有与其他学科信息门户之间实现互操作的可能性。

LII 的资料每周更新,有工作人员负责新资源的追加和资源链接的有效性,据该门户网站的介绍,它们的“死链”从来都没有超过 100 个。同时,鼓励用户根据 LII 的标准推荐新资源,还设立“本周新资源”、“更多新资源”和“上周新资源”栏目,资料更新比较及时。

LII 创造了一个普通用户易于使用的界面,用户可在所有的大小主题中浏览。检索功能也非常强大。基本检索中,可以选择所有字段或在资源题名、主题、描述、URL 字段检索,有词根检索与非词根检索的选择,对输出结果可以选择只显示题名或完整记录。高级检索中还支持布尔逻辑检索。

LII 注意新技术的应用,目前采用快速的、灵活的、功能强大的网页索引系统(Simple Web Indexing System for Humans-Enhanced, 简称为 SWISH-E)。

自1990年面世以来, LII的服务就没有间断过, 可见, LII很重视信息的稳定性与安全问题。

因此, 从资源的收录范围、质量、功能、兼容性、更新与维护、新技术的应用以及资源的可用性方面来看, LII都做得不错。它每月的点击数是120,000, 已被世界范围内的5600多个网站作了链接。

但是, 其工作人员主要是志愿者或兼职人员, 建立和维护的人力仍然不够。而且, 由于开发的时间较早, 只能提供信息的浏览与检索, 而忽略了用户定制等服务。2000年开始的SMETE不只是一个网上的信息存储场所, 而且给用户提供了一个动态的、开放式的数字化学习空间, 用户可在此进行信息浏览与检索、下载、定制个人文件夹、获得适合其需要的资源推送服务、发表评论以及与同行交流等, 这些都是新型的学科信息门户应该具备的。

[参 考 文 献]

- [1] 张晓林. 分布式学科信息门户中网络信息导航系统的规范化建设[J]. 人大复印报刊资料, 2003, (1).
- [2] What is Subject Gateway? [DB/OL]. <http://www.desire.org/html/subjectgateways/subjectgateways.html>, 2003-02-28.
- [3] MANaging the hybrid LIBrary for the Benefit of Users (MALIBU): a hybrid library for humanities models for management and organisation[DB/OL]. <http://www.kcl.ac.uk/humanities/cch/malibu/>, 2003-02-28.
- [4] Engineering E-Library, Sweden[DB/OL]. <http://eels.lub.lu.se/>, 2003-02-28.
- [5] "All" Engineering Resources on the Internet: a Companion Service to EELS[DB/OL]. <http://www.kcl.ac.uk/humanities/cch/malibu/>, 2003-02-28.
- [6] OpenLDAP Software[DB/OL]. <http://www.openldap.org>, 2003-02-28.
- [7] Internet RFC/STD/FYI/BCP Archives. RFC2651[DB/OL]. <http://www.faqs.org/rfc2651.html>, 2003-02-28.
- [8] National Science, Math., Engineering and Technology Education Digital Library (SMETE)[DB/OL]. <http://www.smete.org>, 2003-02-28.
- [9] lli.org Style Manual. Version 2, 2003-02-28.

(责任编辑 涂文迁)

Evaluating Information-organization of Subject-based Information Gateway

HUANG Ru-hua

(School of Information Management, Wuhan University, Wuhan 430072, Hubei, China)

Biography: HUANG Ru-hua (1968-), female, Associate professor, Doctor, School of Information Management, Wuhan University, majoring in information retrieval, networked information organization.

Abstract: The article analyzes characteristics of information-organization of subject-based information gateways (SBIGs), compares famous SBIGs at home and abroad, then puts forwards criteria of SBIGs including coverage ratio of resources in a certain subject, quality and description of resources selected, update and maintenance, interoperation, application of related new technologies, thesauri and safety. Based on the criteria, Librarians' Internet Index (LII) is evaluated.

Key words: subject-based information gateway; networked information organization; criteria