

● 韩 芸

网络信息资源组织方式研究

摘 要 网络为信息资源组织创造了一个前所未有的实验环境和应用情景,许多网络信息资源组织方式应运而生。目前网络信息资源的组织方式有:主题指南方式、自动分类方式、联机编目方式、数据库方式、文件方式、超媒体方式及元数据库方式等。参考文献7。

关键词 网络信息资源 信息资源组织 知识组织方式 元数据

分类号 G253

ABSTRACT The network has created an unprecedented experimental environment and application situation for the organization of information resources, and many patterns of network information resource organization are emerging. In this paper, the author summarizes some patterns, such as online cataloging pattern, hypermedia pattern and metadata pattern. 7 refs.

KEY WORDS Network information resource. Information resource organization. Knowledge organization patterns.

CLASS NUMBER G253

1 网络信息资源组织的含义及网络信息用户的特点

1.1 网络信息资源组织的含义

“网络信息资源”也称“电子信息资源”、“因特网信息资源”、“万维网信息资源”、“联机信息资源”等。迄今为止,对于“网络信息资源”尚未有统一的定义,有人认为“电子信息资源是以电子数据的形式将文字、图像、声音、动画等多种形式的信息存放在光盘等非纸质的载体中,并通过网络通信、计算机和终端等方式再现出来的信息资源”。也有人认为,网络信息资源可以理解为“提供计算机网络可以利用的各种信息资源的总和”等。但以下几点是大家所认可的:第一,网络信息资源是以电子形式存储在光、磁等非纸质的载体中的文字、声音、图像、动画等信息;第二,它利用计算机技术、通信技术及多媒体技术在网络上发布、传递,并且在网络终端得以再现;第三,网络信息资源是指能满足人们需求的那一部分信息,而并非所有在因特网上发布的信息都是网络信息资源。

所谓网络信息资源组织是指人们根据网络信息资源本身的特点或属性,运用各种工具和方法,对网络信息资源进行搜集、加工、整理、排列、组合,使之有利于网络信息资源的存储、传播、检索、利用,以满足人们的网络信息需求的活动过程。网络信息资源组织的对象有:网页、电子邮件、FTP、Gopher、BBS等,其中以网页为主要对象。

1.2 网络信息用户的特点

网络用户的范围十分广泛,用户受教育的程度不同,个体的认识和认识特征、知识储备各不相同,对知识体系的了解也不同;用户的工作岗位和专业不同,对同一概念、同一事物的内涵与外延理解不同、描述的语言不同,对它们在知

识体系中的位置认识也有所不同;各种用户上网的目的和利用信息的种类存在明显差异,由此导致了用户查询和利用信息的视角不同、方法不同、类型不同、深浅程度不同。

但不论网络用户存在多大的差异性,有一点他们是相同的,即他们利用网络信息资源的目的是明确的,他们需要的信息是有序的,并且常常围绕某一个或几个专题。这种信息需求的集中性、有序性和网络信息资源的分散性、无序性之间有着难以调和的矛盾,而网络信息用户的差异性又使得这种矛盾更为突出。因此,只有对网络信息资源进行有针对性地组织,才能使其有序化,方便查询,使网络信息用户真正有效地利用网络信息资源。

2 搜索引擎和虚拟图书馆——网络信息资源组织较成功的模式

搜索引擎主要由巡视软件、数据库和检索代理3部分组成。搜索引擎不是以手工标引,而是以计算机程序(巡视软件)采集和标引网络信息为主的技术,是目前查询检索因特网、新闻小组、gopher、ftp文件、图像、音频和视频信息的最主要的检索工具。网络信息资源的主要特征是以分布式状态存在,因此要以分散的方式获取、检索和处理,其中智能代理是用户进行查询的主要方式。利用搜索引擎,用户只需要在给定的检索框内输入关键词及其组配,或者按照分层类目结构依次逐层选择,就可以得到适合需求的大量网站链接,通过点击超级链接,用户能够访问这些网站,在一定程度上避免了用户浏览网络的盲目性,给用户的信息搜索带来方便。然而,搜索引擎在扩充自己的待建数据库时,往往将大量的商业、文化、娱乐等非学术性站点也囊括其中,这一方面保证了查全率,另一方面却极大地降低了查准

率。科技工作者在查询时,得到的结果往往是成千上万条记录,由于其中包含大量的非学术性站点信息,这在较大程度上增加了信息选择的难度,有的用户甚至感到无所适从。

在这样的背景下,产生了大量面向学术研究的专业搜索引擎——学科虚拟图书馆,如英国的爱丁堡工程学虚拟图书馆(Edinburgh Engineering Virtual Library. <http://www.eevl.ac.uk>)。学科虚拟图书馆的主要功能,是在庞大的因特网信息群中搜索某一学科或某一领域的相关依次网络信息,然后将这些网络信息加以组织,并嵌放在固定的网页上,供用户浏览或查询。对图书馆、情报所等传统信息服务机构来说,因特网的迅速发展一方面给他们提供了丰富的信息资源,他们通过网络提供的信息服务具有信息量大、形式多样、方便快捷的特点,这是传统信息工作所不能比拟的;另一方面,他们又面临着由传统图书馆的服务方式向数字化图书馆服务方式的转化。目前虽然对数字图书馆的说法尚存异议,但不管数字图书馆的具体定义如何,对网络信息资源进行有序化组织,并且提供检索是数字图书馆的重要使命。美国国会图书馆的美利坚记忆(American Memory)、新加坡的丹皮尼斯数字化图书馆(Tanpines Regional Library)以及我国的超星数字图书馆所做出的尝试和取得的成就充分说明了数字图书馆在组织网络信息资源方面发挥的重要作用。

以下将立足于搜索引擎和数字图书馆的工作实践,分析网络信息资源的组织方式。

3 网络信息资源的组织方式

3.1 主题指南方式

主题指南是一种可供检索和查询的等级式主题目录,以超文本链接方式将不同学科、专业、行业和区域的信息按照分类或主题目录的方式组织起来。这些主题目录一般在大类目下分成若干小类目,类目之间按照等级系统排列,然后将人工或巡视软件搜集的网址、主页与主题链接起来,用户通过主题指南中链接的指引,逐层浏览,直至找到所需的信息线索,并通过信息线索直接找到相应的网络信息资源。

利用主题指南、图书馆分类法与主题词表的方法对因特网资源进行组织,有助于为网络资源构造基本结构和框架,使信息分布的无序化以有序化的方式体现。虽然现有的主题指南和分类法在涵盖网络信息的数量和规模上还比较有限,但是反映了网络信息资源组织的一个重要趋势。目前,建立一个涵盖所有网络资源的主题指南似乎还不可能,但如果这种方法在各个学科和各个领域逐步普及,会扩展成为以主题或分类方式组织网络信息资源的体系。

3.2 自动分类方式

自动分类也称为机器分类,是使用机读分类表,通过语义分析和概念分析等手段,自动发现和标引各种文献及网络信息。

目前,国际上对自动分类的研究主要着重于对因特网资

源的自动分类和标引。北欧的 WAIS/万维网项目就是一个自动分类的实验项目,由瑞典伦德大学图书馆和丹麦国立技术图书馆合作进行。WAIS 是提供存取网络数据库的一个通用界面。叙词字段、关键词字段、主题字段以及 WAIS 数据库构建的关键词表中的词,是首先要抽取的对象。

当符号系统与数据库中的词相匹配时,匹配词就被添加到关键词表中并赋予权重。在此基础上,WAIS 数据库中的条目会被加上符号并链接到主题树中。

3.3 联机编目方式

联机编目是数字图书馆网络系统功能中最基本的功能需求。网上各馆在编目时,可利用中央书目数据库中的数据完成自己的作业,同时向其提供各自的馆藏信息。现在的联机目录已开始提供存取文摘和索引数据库以及全文数据库,也能够存取存储在地理位置分散的各个不同机构中的资料目录。

除了简单的查找和显示功能,联机目录还提供打印、下载、用电子邮件发送结果、请求文献传递以及存取电子全文的功能。由于数字图书馆把以前分立的图书馆系统、功能和书目控制传统合并在一起,联机目录在数字图书馆中信息组织的作用也越来越重要。

1991年,OCLC开始启动名为“构建因特网资源目录”(InterCat)的项目,其目的是发动图书馆和高等教育机构密切合作,共同为网络电子资源建立一个可检索的,而且与 USMARC 保持一致的目录,并提供完整的电子资源地址和存取信息。1996年2月,OCLC又推出了 NetFirst 数据库。它采用自动收集、验证技术和传统的标引与制作文摘技术,从各种网络资源中选择并记录一批被判定有长远保留价值的信息。数据库收录范围包括万维网主页 Gopher 服务器、电子论坛、电子期刊和图书馆目录等。NetFirst 可以被视为组织数字化资源的成功例证。

3.4 数据库方式

所谓数据库方式,就是将搜索引擎的巡视软件采集和标引的网络信息资源以固定的记录格式存储,用户通过关键词及其组配查询,就可以找到所需要的信息线索,并通过信息线索直接链接相应的网络信息资源。

数据库内容的扩展主要依靠巡视软件采集信息,或是主页作者向检索系统递交自己的主页地址。目前,几乎所有的检索系统都鼓励作者递交自己的主页地址,检索系统会根据用户递交地址的先后次序,逐次造访,采集并且标引主页内容,并试图与主页上的所有链接建立联系。一些商业性的网址或主页的作者,为了扩大自己的影响,积极地将自己的主页递交给不同的检索系统,使人们通过任何系统,都可以查询到他们的信息,以扩大知名度。这种方式使许多无用的信息和学术性信息混杂在一起,降低了系统检索的准确性和相关性。因而有必要提高巡视软件的设计质量。

3.5 文件方式

用文件方式组织网络信息资源具有简单方便的优点,但随着网络信息量的不断增长和用户对网络信息资源利用的普及,以文件为单位对信息资源共享和传输的方式使得网络负载越来越大;而且当信息结构较为复杂时,文件系统难以实现有效的控制和管理。因此,文件方式只能是网络信息资源组织的辅助形式。

3.6 超媒体方式

这是利用超文本与多媒体技术相结合以组织利用网络信息资源的方式。它将文字、图表、声音、图像、视频等多媒体信息以超文本方式组织起来,人们通过浏览的方式搜寻所需信息,避免了检索语言的复杂性,使人们可通过高度链接的网络在各种信息库中遨游。但这种方式使信息资源的有序化整理和组织存在较大难度。

首先,编目对象难以识别。网络环境下,一系列相关信息源汇集于同一主页下,信息源的著录主体难以确定;而网页的超文本形式更使编目人员难以确定符合用户需求的最佳事项与描述动态文献的固定记录。其次,信息源的不确定性。由于电子地址更改频繁,一旦与物理形态资源相连的著录信息发生变化,编目人员也必须将对发生变化的记录加以注释或编制新的书目记录。

3.7 引文分析法

传统图书馆常常利用引文分析法来组织文献信息,达到最大的效率。在数字化环境下,利用引文分析法来研究网上信息资源的使用情况也开始被越来越多的研究者使用。

有学者记录用户点击某一网页或访问某一图书馆数据库的次数,用户与信息资源之间的连接来评价使用频度。有学者测定了电子期刊在印刷型文献和电子文献中被引用的频度。还有的学者用引文分析法来评价公共图书馆网址,通过测算网址中包括的远程超文本链接的频度,列出公共图书馆的核心网络资源表。

3.8 社会网络分析法

社会网络分析法是既考察结构又研究过程的一种方法和一套技术,重在探讨资源的可得性和参与者之间资源的交换关系。哈丁(J. Hardin)等人把社会网络分析技术移植到数字环境中,引入了一个新概念“赛伯空间构图”(Cyberspacemapping)。其内涵是指当用户存取网络上的文献时,可用联机监控手段来记录所创建的连接。

赛伯空间构图可被用于调查研究国际网络化信息空间的发展、电子社会在这种空间中的形成、超媒体的用途以及使用网络的人采用的导航方法,从中得出的相关研究结果可用于多种目的,如发展馆藏、组织信息、帮助用户识别相关资料或探索知识交换等。

3.9 元数据库方式

元数据就是描述数据的数据,是对特定网络资源的总体管理和记录。它是与对象有关的数据,使潜在用户全面了解对象的存在和特点,并支持各种操作。为了支持分布式网络

环境,必须用元数据来描述资源、方法、数据集和用户。

目前发展最成熟和最完整的元数据格式是美国联邦地理数据委员会的地理元数据标准和美国太空总署制定的数据交换格式。而图书馆界比较熟悉的元数据格式是都柏林核心集。其特点是将各种分散的信息资源统一汇总、加工排序并形成统一的定位、选择和检索系统。它从总体上对网上信息类型的基本特征进行记录和管理,其结果相当于图书馆的总目录,但和传统的印刷目录相比,其记录的内容要丰富得多。它的作用是使用户快速准确地查到分布在各地的信息资源,它可以大幅度减少网上文档传输的次数,提高信息检索效率,而且它对于解决电子信息版本辨识困难、需求转换频繁、网络目录建设费用过高等问题也会有所帮助。

4 网络信息资源组织方式的发展方向

从我国的具体情况来看,网络环境下信息资源的组织问题还是一个有待开拓的新领域。我国已出版的有关因特网的著作、译著和论文中对万维网的调查研究,特别是对万维网的信息组织方式的研究还不够深入和系统,有待于进一步深入研究。网络资源组织方式有3个发展方向。

(1)传统的信息组织方式向知识组织方向发展。网络信息资源的海量性、无序性和动态性对传统信息资源的组织方式提出了尖锐挑战。传统信息组织方式由于自身的局限性,无法有效组织网络资源,这就要求信息组织方式向自动化、智能化的知识组织方式逐渐过渡和发展。

(2)DC元数据成为网络信息资源知识组织的主流方式。DC自身所具有的简洁、互操作性强、国际通用等优势决定了它在网络信息资源组织中的重要地位。在今后的一段时间里,DC将会成为网络信息资源知识组织的主流方式。

(3)多种元数据格式长期并存,实现各种元数据之间的互操作,实现资源共享。目前存在多种元数据格式,这些格式各有优势劣势,各有不同的用途,在不同领域发挥着不同的重要作用,短期内无法互相取代。因此,实现各种元数据格式的互操作就具有重大理论意义和现实意义。一旦实现各种格式之间的互操作,就能充分发挥各种元数据格式的优点,高效组织网络信息资源,实现资源共享。

参考文献

- 1 储节旺,郭春侠.论网络信息组织.情报理论与实践,2000(4)
- 2 闫时.虚拟信息资源及其组织与管理.图书情报工作,2000(4)
- 3 陈树年.搜索引擎及网络信息资源的分类组织.图书情报工作,2000(4)
- 4 刘嘉.网络环境下的学术资源.中国图书馆学报,1999(11)
- 5 付先华.网络信息资源的组织与管理.理工高教研究,2002(4)
- 6 臧国全.论网络信息组织.图书情报知识,2002(3)
- 7 牛文杰.中国数字图书馆建设的发展现状.信息技术,2002(2)

韩芸 北京大学信息管理系在读博士研究生。通讯地址:北京中国人民大学图书馆教学研究部。邮编 100872。

(来稿时间:2003-04-10)