

●过仕明 王丽伟

主题门户理论研究

摘 要 主题门户是因特网信息资源发现的一种工具,它相当于因特网上的学术或专门图书馆。主题门户的特点有:信息资源选择,信息资源集合的维护,信息资源描述,主题分类。主题门户能改变网络有效信息不足的现状,增强网络对用户的亲和力。主题门户相对于搜索引擎来说,对信息的揭示更专业、更深入。表1。参考文献13。

关键词 因特网 主题门户 网络信息资源 信息揭示

分类号 G203

ABSTRACT Subject portal is a tool for information discovery on the Internet, and can be regarded as an academic or special library on the Internet. In this paper, the authors discuss the characteristics of subject portals in detail. 1 tab. 13 refs.

KEY WORDS Internet. Subject portal. Network information resource. Information description.

CLASS NUMBER G203

主题门户是因特网信息资源发现的一种工具。然而到目前为止,主题门户还没有明确的定义。究其原因,可以从因特网的动态性得到答案:网络上类似主题门户的服务正快速增加;现有的主题门户服务范围、内容、方法和特点上在不断地变化。因此很难做到定义上的准而全。本文试图在以往的研究基础上对主题门户的理论做一些探索。

1 主题门户的定义

许多早期的门户起源于搜索引擎和有助于用户获取因特网内容的软件程序。今天,搜索引擎仍是门户的组成部分,而且已成为全部门户策略的一部分。搜索引擎可以帮助用户通过检索获得信息。为了减少搜索时间和帮助那些没有经验的用户,一些搜索引擎加入了类目,通过预先设置的主题进行网址和文件的过滤,例如体育、气象、旅游、财经、新闻、文化等。在结合了虚拟交流、实时聊天的功能,以及设置个性化搜索引擎界面、获得专门和商业内容的性能后,这个新概念被称作门户。目前,门户可分为水平门户和垂直门户。水平门户提供的信息服务虽然覆盖面很广,但大多是罗列传统媒体信息,附带介绍大众关心的消息。用户花费大量时间和金钱在水平门户上希望获得少而精的“所见即所得”信息,但用户一点击浏览后,却没有发现符合其需求的信息。水平门户最为典型的一个例子是 Yahoo——来源极其广泛,但内容却很浅。随着网民的多极化群体出现,人们寻求的信息更多地趋向专业化和个性化。网络的发展

客观上需要一批能够满足特定领域人群、特定需求的专业化用户,垂直门户就是基于此目的而产生的。大学或图书馆更可能建设垂直门户,集中于特定的用户群,提供更专门的服务,这就是主题门户,或称学科信息门户。

不同的机构、学者对主题门户给出了不同的定义。

澳大利亚主题门户论坛对主题门户定义是:“基于网络获取高质量及评价过的资源的一套机制,用来支持特定主题的研究。”主题门户是通过某个入口获得的服务,通过开放的标准协议(例如 Whois++, LDAP, Z39.50, Harvest Broker)来实现无缝链接。终端用户看到的是基于网络的简单的界面,而实际上其组织的结构化程度比看到的更高^[1]。

IMesh 是英国的一个为期3年的研究项目,始于1999年,目标是支持主题门户提供者和相关方的交流与合作,为促进主题门户的发展提供一个交换想法和技术的开放论坛。IMesh Toolkit 项目给出了下面的定义:“主题门户是围绕特定主题对在线资源提供搜索和游览入口的网站。主题门户的资源描述通常是手工创建的,而不是通过自动的程序产生。因为资源的数目由手工产生,通常优于那些从传统的网络搜索引擎获得的数目。”^[2]

DESIRE 是于1998年到2000年在欧洲执行的一个项目,包括4个国家10个参与者,解决主题门户开发和利用的不同方面问题。它的主要的并且目前仍很实用的一个产品是建设门户服务的详细手册。

DESIRE项目提出主题门户的定义是:“因特网上的选择性主题门户的特点是质量控制。资源选择和描述的核心工作依赖于熟练的人工输入(由图书馆员、学者和专家),不是自动化的。”^[3]

Traugott Koch认为:主题门户是支持系统性资源发现的因特网服务。它们提供对那些主要可以通过因特网获得的资源(文件、集合、网址或服务)的链接^[4]。主题门户的服务是基于资源描述的。通过主题结构来浏览资源是一个重要的特点。

IMesh Toolkit项目的定义虽然强调了人工的资源描述,却没有明确资源是否经过筛选,其质量是否受到控制。DESIRE给出的定义忽略了对象资源,没有明确主题门户所提供的是何种资源。Traugott Koch给出的定义有几点值得推敲:未说明主题门户的质量控制特征,这就容易与一般的网络服务,如因特网资源目录混淆;没有明确主题门户的人工资源描述过程。澳大利亚学科信息门户论坛定义虽然简洁,却隐含了人工资源描述的过程,相对来说比较完整准确。

Traugott Koch还进一步深化了主题门户的定义,将它分为一般的主题门户和质量控制主题门户:描述较少、主题结构肤浅的链接列表,叫做主题门户(比较上述);质量控制水平较高,描述和结构丰富的主题服务,叫做质量控制主题门户^[5]。

质量控制门户是使用一整套质量控制机制来支持系统资源发现的因特网服务。使用相当多的人工来确保所选资源满足质量标准,并利用标准元数据对这些资源进行详细描述,进行定期的检查和更新来确保较好的资源集合管理。主要的目标是通过使用可控词对资源进行索引,以及对高级搜索和浏览提供深度的分类结构,来提供一个高质量的主题入口。Kirriemuir的调查研究中使用了两个规则来选择质量控制主题门户:门户的内容至少有一部分是英文的,门户服务应提供相当数量的资源——至少是几百个。

Wheatley认为,主题门户在近期发展中,让人失望的事情之一就是没有广泛使用可控词汇提供资源的入口^[6]。因此,Traugott Koch提出的质量控制主题门户应该是目前门户发展的方向。

了解主题门户的定义,有助于人们辨析几个相似的概念:

“基于主题(学科)的信息门户”(SBIG),是由欧盟资助的DESIRE项目提出,基本上与主题门户同义。

“资源发现系统”,其代表的概念比较广泛,包括

许多其他的发现工具,例如网络索引、搜索服务、地方的数据库或图书馆的OPACs。

“主题服务”,当一项服务将许多其他的不属于门户的特点和服务加到主题门户上时,可以使用“主题服务”来代替主题门户^[7]。

“因特网资源目录”(IRC),是通过结构化的网络服务获取的因特网资源描述的数据库。它只是门户提供的一项服务^[8]。

那些纯粹是当地的资源目录或索引不能叫做门户,包括或多或少因特网资源的图书馆目录,尽管质量控制水平较高,但叫“主题门户”也不恰当,因为它们主要列出的是当地的或物理的资源,而不是因特网上的分布式资源。

2 主题门户的特点

(1)信息资源选择。主题门户信息资源是由专家根据明确的甚至是严格的选择标准精心挑选的。一般根据信息资源的质量、权威性、可获得性、时效性和主题相关性来选择,其他的标准也可能包括语言或地域范围。例如:MedHist(<http://medhist.ac.uk/>)明确规定:信息资源在列入门户之前必须先行评价,如果网站的目标对象不符合MedHist的目标对象要求,就不能收录;如果对信息资源的准确性和适宜性有任何质疑也不能收录;同时信息资源不能是过时的并可以定期维护,除非是惟一的并且特别重要的信息资源;能够获得、具有可用性。

(2)信息资源集合的维护。信息资源的集合需要定期维护,包括删除那些不适当的信息资源、被取代的信息资源或含有错误款目的信息资源。可能也需要对链接定期检查。这些工作可以自动完成,也可以手工完成。例如,检查链接可使用ROADS link checker, Xenu, LinkBot,或者自己研发的工具。

(3)信息资源描述。选择的信息资源由人工进行描述,采用一定的标准元数据编目。元数据使得信息资源的发现和定位变得容易,也可以促进结构化检索。描述包括信息资源的内容、作者、出版者或出版日期等信息。例如,英国的资源发现网络RDN使用了IAFA/ROADS文档和SERVICE模版、都柏林核心集,瑞典的工程电子图书馆EELS使用了IAFA模版标准(来自ROADS系统),可转换为都柏林核心集。

(4)主题分类。人工使用主题分类对所有的信息资源进行索引,可以促进主题浏览,例如RDN和EELS都是用了国际主体分类体系。

不同的主题门户还有各自的一些特点,例如:

(1)SOSIS(社会科学信息门户)提供了包含社会科学术语的主题词表,并允许用户对位于某一特定地理位置的资源进行搜索,例如世界范围、欧洲或英国等^[9]。EEVL(爱丁堡工程虚拟图书馆)允许用户对某些资源类型进行搜索,例如电子期刊或会议资料等^[10]。

(2)基于主题的服务。一些主题门户提供了相关的服务。例如:Biz/ed(学生、教师和讲师的商业经济信息门户)包括了大量的商业和经济信息^[11];OMNI,医学与卫生主题门户,允许用户购买与健康相关的CD-ROMS^[12]。

(3)个性化信息定制。中国科学院国家科学数字图书馆的学科信息门户群的宗旨之一是逐步支持开放式集成定制。

3 主题门户的性质

在某种意义上,主题门户相当于因特网上的学术或者专门图书馆,它具有如下性质。

(1)主题门户是促进因特网信息资源发现的服务。主题门户的信息资源不是对网上信息资源的简单罗列。它以用户为中心,由学科专家和信息专家评价并描述,有利于因特网信息资源的发现,具有信息增殖的特性。

(2)主题门户是网络信息资源的导航,是网上的

三次信息,指向网上的一次信息(文件、超文本等)和二次信息(搜索引擎、指示数据库等)。建立主题门户的目的是为用户提供便捷的导航,它具有严格的质量控制,针对目标群体,对入选信息资源进行描述甚至评述,帮助用户快速、准确地发现高质量的信息。用户只要来到相应的主题门户,就可以在它的指导下找到需要的信息。

(3)主题门户增加了网络对用户的“亲和力”,增强网络的实用性和服务性。主题门户以新颖的、有特色的服务切入市场,可以吸引更多的、潜在的用户群体。主题门户专注于信息的专深性,更体现以人为本的理念,更吸引用户的注意力。

(4)主题门户是对信息构建理论的深化。信息构建的核心理念是关注用户,以人为本。信息构建是信息用户、信息内容与信息组织三者的交集。主题门户网站的建设充分考虑了信息有效传导的要求,立体地呈现出特定主题的各类信息资源(电子期刊、数据库、知识库、讨论组等),符合特定用户的需求,提高了资源获取的效率。

虽然门户是由搜索引擎发展而来,但主题门户与搜索引擎在性质上并不相同(表1)。主题门户对信息的揭示更专业更深入;而搜索引擎相对广泛、肤浅。而且某些“隐藏网”内的信息(“隐藏网”指的是搜索引擎找不到的并且很少包括在主题目录中的网页),只有主题门户能够揭示。

表1 主题门户与搜索引擎的区别

搜索引擎	主题门户
资源集由自动程序创建	资源集由人工创建
资源集无目标用户	资源集以目标用户为中心
目标是无遗漏地对资源进行索引	目标是建立高质量的资源目录
提供自动生成的元数据	提供人工生成的元数据
终端用户得到的结果质量较差	终端用户得到的结果质量较高
索引的对象是网页,对资源的描述不连贯	索引的对象是资源,对资源的描述连贯
收录的资源规模较大	收录的资源规模较大

对主题门户来说,人工地选择和创建记录是必需的,与搜索引擎相比,主题门户收录的资源规模较小。人工语义加工的缓慢耗时和用户需要更多门户资源内容的矛盾是目前门户的一个主要问题。研究人员正在寻求快速创建大规模门户的方法。

4 主题门户入口

主题门户可以通过发现因特网信息资源的普通方法来确定,例如搜索引擎(Google);也可以通过电子讨论列表和讨论组提供的信息进入;或者通过个人的建议进入^[13]。目前,许多图书馆和信息服务机构,不再只是列出其领域范围的所有有用的因特网资源,

而仅仅列出主题门户。

下面介绍几个主题门户的列表:

(1) Heriot Watt 大学的 PINAKES (主题导航簿) 门户列表 (<http://www.hw.ac.uk/libWWW/irn/pinakes/pinakes.html>)。PINAKES 提供了吸引眼球的主题门户列表,但对各个主题门户的评价信息很少。

(2) Clearinghouse (<http://www.clearinghouse.net>),这是一个“通往门户的门户”。资源指南按照下列标准对各个门户进行描述并对其质量进行“评分”:资源描述;资源评价;指南设计;组织方案和指南的元信息。

(3) 中国科学院国家科学数字图书馆 (www.csdn.ac.cn),目前已经建成了6个主题门户(物理、数学、化学、生命科学、资源和环境科学、图书情报),是国内权威、可靠的主题门户体系。

5 结束语

主题门户是图书馆和信息专业人员对资源发现问题问题的一个特别贡献。但从目前情况看,主题门户的利用范围较小,还未被网络用户真正重视起来。通过主题门户促进网络资源的利用,既是事关我国互联网信息服务发展的重大课题,也是未来信息工作者的一项重要任务。

参考文献

- 1 澳大利亚主题门户论坛. <http://www.nla.gov.au/initiatives/sg/servicetypes.html>

- 2 Imesh 门户. <http://www.imesh.org>
- 3 DESIRE 项目网站. see <http://www.desire.org>
- 4,5,7 Traugott Koch. Quality-controlled subject gateways: definitions, typologies, empirical overview, Manuscript of the article published in the Subject gateways special issue of "Online Information Review" Vol. 24: 1, Feb 2000. Copyright: MCB Univ. Press
- 6 Wheatley, A. (2000). Subject trees on the Internet: A new role for bibliographical classification. In A. R. Thomas, & J. R. Shearer (Eds.), Internet searching and indexing: The subject approach (pp. 115 - 141). New York: Haworth Press
- 8 英国的资源发现网. <http://www.rdn.ac.uk/publications/terminology/>
- 9 SOSIG. 社会科学信息门户. <http://www.sosing.ac.uk/>
- 10 EEVL. 爱丁堡工程虚拟图书馆. EEVL allows you to search for resources of a particular type e. g. mailing list. <http://www.eevl.ac.uk/search.html>
- 11 Biz/ed. 学生、教师和讲师的商业经济信息门户. <http://www.bized.ac.uk/>
- 12 OMNI. 医学网络信息. <http://omni.ac.uk/>
- 13 David Bawden, Lyn Robinson. Internet subject gateways revisited. International Journal of Information Management, 22(2002)

过仕明 哈尔滨师范大学图书馆馆长, 吉林大学情报学博士研究生。通信地址: 哈尔滨市。邮编 150080。

王丽伟 吉林大学公共卫生学院医学医药信息管理学助教, 吉林大学情报学专业在职硕士研究生。通信地址: 长春市。邮编 130000。 (来稿时间: 2004-10-15)

(上接第44页) 完善和强化各校图书馆的网络功能,立足于进一步提高全区高校图书馆的现代管理服务水平,不断创新服务手段,不断建设,形成地域性、系统性、规模化和特色化。

3.7 培养一支高素质的人员队伍

人是特色化建设的执行者,是特色化建设成败的关键因素。高校图书馆面对的是高素质的服务对象,服务人员也必须具备高素质。特色化建设要求服务人员具有良好的敬业、开拓创新精神,主动服务观念;有良好的专业知识背景;有较强的计算机和网络技术应用能力;有较高的外语水平和民族语言的表达能力。只有这样才能保证特色化建设的成功,保证特色服务的有效开展。

参考文献

- 1 于建荣,伍宗韶. 专业图书馆的服务创新. 图书情报工作, 2003(4)
- 2 龙世谱. 论我国西部图书馆的特色化建设. 图书情报工作, 2003(1)
- 3 石丽珍. 关于优质服务的几点思考. 中国图书馆学报, 2003(4)
- 4 顾英. 高校图书馆学科特色数据库体系的构建和实施. 上海高校图书馆情报工作研究, 2003(2)
- 5 阿拉坦仓. 人本管理——大学图书馆管理的瓶颈. 大学图书馆学报, 2004(5)

阿拉坦仓 内蒙古大学图书馆馆长,教授。通信地址: 呼和浩特市大学西路235号。邮编 010021。

(来稿时间: 2004-11-19)