

图书馆个性化主动服务实践及其信息组织

郭家义

(中国科学院文献情报中心,北京,100080)

[摘要] 本文论述了图书馆个性化主动服务的各种实践活动的信息组织方法,并对图书馆提供个性化服务及其信息组织存在的问题及发展方向等进行了探讨。

[关键词] 数字图书馆 个性化 服务 信息组织

[中图分类号] G252 **[文献标识码]** A **[文章编号]** 1003-2797(2004)01-0070-04

1 图书馆个性化主动服务实践及其信息组织概述

1.1 MyLibrary

美国图书馆学会 ALA(American Library Association)^[1] 下属的来自图书馆和信息技术领域的专家小组 LITA(Library and Information Technology Association) 每年或每季度都要对图书馆技术的发展作出预测。1999年1月,他们预测的图书馆研究的七大趋势是:定制与个性化、网络资源评价、人文因素、Co-opt Existing 技术、家庭学者、认证和权限管理、Submerging 技术^[2]。定制和个性化服务被列为七大趋势之首,从而可见其重要性。MyLibrary 的不断涌现和发展也印证了这一点。据笔者调查,目前美国大多数大学图书馆都建有自己的 MyLibrary。我国图书馆对定制与个性化的研究和实践起步比较晚,目前只有中科院文献情报中心、国家数字图书馆、浙江大学图书馆等少数几家图书馆开展了此项服务。可喜的是,国内一些专家学者如张晓林、赵继海、余锦凤等在个性化定制方面做了有益的探讨和研究,并取得了一定的研究成果^[3~5]。

MyLibrary 进行信息组织时需要考虑的因素包括:①服务对象与用户类型;②定制的内容及资源;③什么内容最为重要;④包含的资源种类及多少;⑤通过 e-mail 发送什么信息;⑥各标题款目要有一定的控制,力求形成一个分类系统。MyLibrary 更多地依赖数据库技术和动态网页技术,它从本质上并没有改变图书馆原有的信息组织方式。它通过一定的机制,为用户创造对图书馆资源进行重组(扩大或缩小资源范围)或选择的机会,实现某种程度的个性化。同时,MyLibrary 提供的是图书馆的资源与服务的链接,没有实现资源与服务的整合。目前已经在探索进一步提高 MyLibrary 的个性化服务水平,将 MyLibrary 构筑在集成的图书馆资源与服务的基础上,为用户提供一站式检索与一站式服务,

这种建设方式将对 MyLibrary 的信息组织方式提出更高的要求。

1.2 Active learning 个性化工具

为用户建立 Active learning 个性化工具,是图书馆为适应网络环境下的教育服务所进行的一种新尝试。Active learning 个性化工具作为传统检索与浏览工具的扩展,为用户进行个性化学习提供了有效的支持。其主要功能有主动消化、信息搜集、信息查找。主动消化是用户可以有效地使用各种类型的信息资源(文本、图像、多媒体等),它主要包括主动阅读、主动观看和主动倾听^[6]。Active learning 个性化工具的信息组织方式有:

(1)浏览结果总结。将多媒体信息资源以便于用户注释、分割和整合的机制存储,并为用户提供相应的工具,用户利用这些工具可以很方便地将图书馆资源组织到自己的个性化空间中。

(2)资源分类。包括个性化图书、个性化论文、个性化 Audio/Video 和个性化资料等。

(3)拷贝资源的组织。当用户对资源进行拷贝,建立自己的个性化资源时,系统自动为用户所选定的资源建立指针,指向用户个人文档。这种方式不进行资源的物理拷贝,提高效率,节省了存储空间。

Champa Jayawardana^[7]等在建立 Active learning 个性化工具时,尝试对各种信息建立特定的信息组织方式,并对不同媒体形式进行不同的映射。具体方法如下:

(1)文本数据的组织。文本片断有三个属性,即文档标识、起始偏移地址和最终偏移地址。同时每一个文本片断标识包含物理声明和逻辑声明。文档标识的物理声明是资源对象的 URL 或 URI。逻辑地址包含可以有效搜索文档物理位置的有用信息,例如 ISDN 号。起始偏移地址和最终偏移地址用于检索文档的某一片断。

(2) 图像数据的组织。图像的左上角的地址作为起始偏移地址,右下角的地址作为最终偏移地址。

(3) Video 数据的组织。Video 数据有三个属性,即资源标识、开始 Frame 地址和最终 Frame 地址。资源标识包括文档中资源的物理地址。开始 Frame 地址和最终 Frame 地址可以通过某种匹配函数计算出来。

1.3 Alerting system

Alerting system(Awareness Services, New Material Services)系统是依据用户的兴趣为用户提供图书馆最新资源的个性化主动服务系统。笔者在文献[8]中对这种系统的起源、设计以及需要注意的问题做了一定程度的探讨。目前,大多数商业网站提供分类列表供用户选择服务类型和内容,并允许用户设定服务的内容、方式和其它要求。在图书馆方面,英国图书馆的 Zetoc^[9]对此进行了有益的探讨。

Zetoc 是英国图书馆与英国高等和远程教育联合信息系统委员会(Joint Information Systems Committee for UK Higher and Further Education)共同建立的资源服务系统,它是分布式国家电子资源发展项目(Distributed National Electronic Resource Development Program^[10])的一部分。通过电子邮件,Alerting 服务为用户提供英国图书馆新到期刊的目录列表。Zetoc 目前的版本是 Z39.50 Bath Profile,可以提供使用 XML 格式编码的 Dublin Core 记录,以响应 Z39.50 检索请求。Zetoc 的加强版的原型正在建设之中,加强版将使用 XML 仓库保存数据,使用 Dublin Core 作为基本元数据 schema。Zetoc 提供了在几种开放的标准数据格式间互操作的实践,主要有 Zetoc 记录与 Dublin Core 的映射、Zetoc Dublin Core 与 Z39.50 的映射、Zetoc Dublin Core 与 OpenURL 的映射。^[11]

(1) Zetoc 记录与 Dublin Core 的映射。Zetoc 记录的大多数字段与 Dublin Core 的元素相对应,包括文章题目、作者、国会图书馆主题编码、Dewey 分类号、出版年等。会议文献的某些字段中包含标识文献在英国图书馆的馆藏位置的信息,建立这些数据与 Dublin Core 的映射时需要使用 Dublin Core Qualifiers^[12],同时需要使用 Zetoc 特定的 Qualifiers。在某些情况下,需要使用 Zetoc 名字空间的子元素来表示文章不同部分的作者。

(2) Zetoc Dublin Core 与 Z39.50 的映射: Zetoc 除了有 Web 检索功能外,还提供 Z39.50 检索界面。它允许通过 Z39.50 Bib-1 属性集(Bib-1 At-

tribute Set)^[13]进行检索,并返回 SUTRS 格式、GRS-1 格式或者简单标签格式的信息。为了适应 Bath Profile 标准,Zetoc 也返回 Dublin Core。该系统中分别对 Bib-1 和 GRS-1 建立了映射表,直接进行数据映射转换。①与 Bib-1 的映射:由 Cheshire 软件将由 XML 数据文件建立的索引映射到 Z39.50 Bib-1 属性。这允许 Z39.50 客户端检索 Zetoc 记录的特定领域。②与 GRS-1 的映射:GRS-1 (Generic Record Syntax)是 Z39.50 协议定义的记录检索词法。

(3) Zetoc Dublin Core 与 OpenURL 的映射。Zetoc 系统将检索结果与英国图书馆的馆藏紧密结合起来,用户通过 Zetoc 检索到的记录会通过映射直接链接到该馆。Zetoc 系统使用 OpenURL 协议的“Object-Metadata-Zone”建立从检索结果到该馆馆藏记录的链接。

1.4 个性化信息环境(PIE)

数字资源的有效利用一直是图书馆努力解决的问题之一。某些图书馆为此进行了个性化信息环境(Personal Information Environment PIE)建设的实践。主要项目有我国台湾交通大学图书馆 PIE^[14]、网络环境下图书馆复合电子资源访问和传递的 HeadLine^[15],以及 Virginia 大学和 North Carolina 大学信息和图书馆科学学院的 NCSTRL^[16]等。

HeadLine 的 PIE 从其功能和信息组织方式来看,仍是 MyLibrary。虽然它通过评估提出了改进系统的想法,不过从目前情况来看,该项目已经处于停顿状态,没有新的改进。我国台湾交通大学图书馆的个性化信息环境的基础也是 MyLibrary,但该系统已开始对用户信息进行挖掘,并依用户兴趣进行信息组织,相对 MyLibrary 系统又发展了一步。比较系统提出、研究、建设个性化信息环境的是 Virginia 大学和 North Carolina 大学信息和图书馆科学学院,其建设的原型系统是 NCSTRL。

NCSTRL 的主要思想是建立持久的、用户可定制的、私有的或可共享的、提供有效和有效率的检索和安全的个人信息环境。该系统涉及的研究领域包括分布式检索、信息收集和元数据管理。该系统实践中将个性化信息环境分为个性化收藏和虚拟存储。个性化收藏是用户指定的信息资源,虚拟存储是对信息资源的统一封装。封装信息中包括文档计数、作者列表、站点名称、资源类型、资源总结信息。资源可以是个人,也可以是复合的。用户通过定制建立个性化收藏。目前,该系统已经能够实现个性化资源的建设和分布检索。

2 个性化服务与知识组织探讨

2.1 PIE的评估

英国高等教育联合信息系统委员会建立了个性化服务系统——PIE^[17],并分两个阶段对该系统进行了评估^[18~19]。

(1)评估内容。包括用户数量、用户类型、用户国籍、PC机类型、使用次数、使用图书馆资源情况(不是通过PIE使用)、通过PIE使用资源的情况、用户使用PIE特定页面的情况、使用频率、个性化和检索功能的使用、界面的可用性、PIE工具的可用性、系统响应时间、用户需要的其它功能。

(2)评估方法。采用用户访问日志分析、用户调查表。

(3)评估结果。包括以下内容:①PIE的频繁利用率低;②普通用户会使用PIE,但次数很少;③图书馆工作人员、进行相似工程的人员会使用PIE,他们的目的是了解PIE的功能和特点,并与特定系统进行比较;④经过培训的用户利用PIE更频繁;⑤用户认为需要改善系统的用户界面设计和提高系统响应时间;⑥用户认为PIE应与图书馆现有的系统整合起来。

(4)用户的需要。包括以下内容:①对图书馆资源的整合检索;②检索结果合并;③对PIE内资源的无缝访问(一次登录PIE就可以利用其资源,无需多次输入用户名和密码);④current awareness服务;⑤文档传递服务;⑥从PIE检索界面使用网络搜索引擎;⑦检索结果以书目格式输出;⑧将链接导入用户浏览器的个人收藏夹中。

2.2 Cyber Dialogue的调查^[20]

Cyber Dialogue是eCRM供应商,它主要致力于战略信息和分析存取、定制研究和数据挖掘服务、Internet市场软件和技术。它对商业服务提供技术、系统和用户挖掘支持。它在2000年和2001年对用户隐私和个性化进行了分析与调查,得到了一些有用的数据。①2000年利用个性化的用户达到2820万,比过去两年增加了十倍;②只有58%的用户理解cookies在内的个性化技术;③用户可以提供包括姓名、教育水平、年龄、习惯和特定兴趣以及家庭收入等信息;④用户隐私信息可能被滥用。

根据调查结果,Cyber Dialogue对网站提供个性化服务提出如下建议:①建立与用户的长期关系而非短期行为;②在网站上要有一个清晰明确的隐私声明;③确保收集用户授权的信息;④用户将选择那些与之建立良好关系的站点;⑤只收集与个性化

过程相关的用户信息;⑥听取用户的意见与建议;⑦防止合作者对用户信息的滥用;⑧如果不注意保护用户的隐私将失去用户;⑨获得良好的用户关系有助于服务水平的提高。

2.3 关于个性化服务的思考

(1)图书馆的反思。Cyber Dialogue的调查表明,接受个性化服务的商业网站用户在过去的时间内大幅增加;但PIE的评估结果则表明,利用图书馆个性化服务的用户的数量并没有令人振奋的结果。这可能需要反思图书馆在个性化服务中的服务理念、服务方式、服务框架等。笔者认为,一个图书馆的个性化服务的出发点决定了其服务的效率,全面改进图书馆的个性服务效率需要图书馆从什么是个人信息环境、什么是图书馆个性化服务、图书馆在建立个人信息环境中的地位和作用等方面进行思考,建立一种全新的个性化服务框架。

(2)隐私保护将成为个性化服务发展中的关键问题。两项调查表明,首先,用户会担心自己的个人信息可能被滥用。其次,用户对个性化服务中采用的技术并没有清晰的了解。个性化服务方(图书馆、商业网站或其它信息服务商)需要采取充分措施保护用户的个人信息。正如Cyber Dialogue建议的,要与用户建立长期的合作关系,而非短期行为。同时个性化服务方需要加大宣传力度,使得用户了解个性化服务可以为其带来的益处。

(3)用户的需求是个性化服务的出发点。PIE的调查结果表明,目前,图书馆的个性化服务并没有充分考虑用户的需求,仍停留在以图书馆为中心的信息资源建设与服务模式上,没有真正建立以用户为中心的信息资源建设和服务模式。图书馆个性化服务的出发点需要跟随用户的需求而不断变化,及时跟踪这种变化的前提是图书馆要建立一套个性化服务机制,而非资源的简单收藏。

(4)个性化服务质量有可能决定个性化服务的前途。目前,个性化服务,如最新资源通报、定题服务等,为用户提供的信息资源仍存在大量的与用户信息需求不相关的信息。这既有信息服务方的原因,也有信息需求方的问题。从信息需求方来说,信息需求方不能很好地描述自己的信息需求,同时他们的信息需求又是不断变化的。从信息服务方来看,信息服务方不能很好地利用用户需求信息,不能对用户信息需求进行深入理解。从某种意义上讲,个性化服务的质量将决定个性化服务的前途。

2.4 个性化服务中的信息组织

2.4.1 个性化服务中信息组织存在的问题

(1)没有有效的个性化信息组织方式。信息组织是图书馆界非常重视的研究课题之一,但目前还没有全面的理论研究和探讨,所有的个性化服务中的信息组织仍是利用普通环境下的信息组织方式,有效的个性化服务信息组织方式还没有形成。

(2)没有有效的个性化服务集成模式。个性化服务是在原有服务集成的基础上进行,还是另辟蹊径,是值得研究的问题。目前存在多种个性化服务模式,用户需要适应不同的入口和不同的使用方式,因而造成了用户利用图书馆个性化服务的障碍。

(3)个性化信息组织理论仍需加强。个性化信息组织理论是一个全新的研究领域,系统地建立个性化服务的信息组织理论是图书馆有效提供个性化服务的基础。

(4)强调技术的重要、忽视理论的研究。毋庸置疑,各种技术的出现为图书馆更为有效地为用户提供服务提供了新的基础和新的渠道,但如果忽略对个性化理论本身的研究,不能足够了解用户的“个性化”需求,就会导致资源与服务利用率低下。

(5)没有从用户角度进行信息组织。虽然图书馆试图为用户提供符合个人要求的信息资源服务,但从严格意义上讲,图书馆仍然是从图书馆的资源角度进行信息组织,没有从用户个人信息环境角度考虑图书馆能为用户提供什么信息资源。

2.4.2 个性化服务信息组织的发展方向

(1)加强理论研究。从目前的文献调查可以发现,个性化服务中的信息组织还没有形成理论体系。而个性化信息组织理论是个性化服务的关键问题。目前个性化服务在知识组织方面所做的探讨,为建立系统的个性化信息组织理论提供了基础。

(2)从资源的简单组织到强调个性化信息重组。个性化服务的初期是对资源的简单组织,提供的是图书馆原有的信息组织体系。研究表明,传统的信息组织方式已经不适应个性化服务的要求,新的个性化信息组织方式是个性化服务发展的必然要求。

(3)从以图书馆为中心进行信息组织到以用户为中心进行信息组织。PIE评价已经在考虑用户的要求,但遗憾的是该系统没有作进一步的改进。可喜的是,它已经认识到用户的需求是图书馆提供个性化服务的基础。提供全面的个性化服务,需要图书馆从用户信息需求的整个环境来考虑信息组织,信息组织的出发点是用户,而非图书馆。

(4)从链接的简单集成到资源原文提供。随着

技术的发展,用户的信息需求在不断变化。用户已经越来越不满足于链接资源的简单提供,而是希望直接链接到图书馆资源,进一步获得信息的原文资料,而非题录或文摘。

(5)从资源集成到服务集成。用户需要将其所需的图书馆资源与服务集中在用户指定的存储系统中,并通过一次利用个性化服务就可获取所有的个性化资源和服务。同时,用户无须面对和适应不同的个性化资源和个性化服务系统。服务集成提高了个性化服务对用户的友好性与信息利用的效率。

参考文献

- 1 <http://www.ala.org/>
- 2 http://www.ala.org/Content/NavigationMenu/LITA/LITA_Resources_and_Services/Top_Technology_Trends/Midwinter_1999.htm
- 3 张晓林等.基于Web的个性化信息服务机制.现代图书馆情报技术,2001(1)
- 4 赵继海.论数字图书馆个性化定制服务.中国图书馆学报,2001(3)
- 5 余锦凤.基于WEB的数字图书馆定制服务系统.<http://www.npopss-cn.gov.cn/planning/xm2000/tsg.html>
- 6,7 <http://wings.buffalo.edu/publications/mcjrnl/v8n1/active.pdf>
- 8 郭家义.个性化最新资料快报服务系统设计及相关问题研究.现代图书馆情报技术,2003(3)
- 9 <http://zetoc.mimas.ac.uk>
- 10 The UK Distributed National Electronic Resource (DNER). <http://www.jisc.ac.uk/dner/>
- 11 <http://jodi.ecs.soton.ac.uk/Articles/v02/i02/Apps/apps-v2.pdf>
- 12 Dublin Core Qualifiers. <http://www.dublincore.org/documents/dcmesqualifiers/>
- 13 The Z39.50 Bib-1 Attribute Set. <http://lcweb.loc.gov/z3950/agency/defns/bib1.html>
- 14 柯皓仁.数字图书馆现状与发展. <http://www.lib.nctu.edu/>
- 15,17 <http://www.headline.ac.uk/>
- 16 <http://www.cs.virginia.edu/~cyberia>
- 18 <http://www.dlib.org/dlib/march01/gambles/03gambles.html>
- 19 <http://www.headline.ac.uk/public/ph2guestreport.pdf>
- 20 <http://www.cyberdialogue.com/library/pdfs/wp-cd-2000-privacy.pdf>

作者简介:郭家义,男,1969年生,馆员,在职博士研究生。

(收稿日期:2003-09-10)