

黄碧云 方平

## WWW 联机公共检索目录研究进展

**摘要** WebPAC 即 WWW 联机公共检索目录。它具有图形用户界面;可显示全文、图像和多媒体信息;可提供用户界面最大弹性,可检索远程数据库,能链接全文。从改善 WebPAC 书目记录内容、书目记录显示和用户界面以及提高 WebPAC 检索功能三个方面,论述了 WebPAC 的研究进展。参考文献 30。

**关键词** WWW 联机公共检索目录 书目记录 检索功能

**分类号** G354.2

**ABSTRACT** WebPAC, or WWW public access catalog, has the following characteristics: graphic user interface, full-text and multimedia display, flexible user interface, remote database access and full-text linking. In this paper, the authors summarize recent researches in this field. 30 refs.

**KEY WORDS** WebPAC. Bibliographical records. Search functions.

**CLASS NUMBER** G354.2

80 年代以后联机公共检索目录(Online Public Access Catalog, OPAC)开始出现,起初可视为卡片目录的机读版本<sup>[1,2]</sup>。80 年代中期出现第二代 OPAC,功能有很大提高,如提供布尔逻辑检索、邻近检索、截词检索和关键词检索,可选择不同的显示格式,并可提供联机帮助<sup>[3,4]</sup>。

90 年代后期,由于 Internet 和客户机/服务器模式的发展,WWW 联机公共检索目录(Web-based OPAC,即 WebPAC)开始出现<sup>[5]</sup>。WebPAC 是指利用 Web 技术的 OPAC。WebPAC 除具有第二代 OPAC 的特点,还采用 Internet 用户所熟悉的图形用户界面,用户无须记忆检索指令,掌握复杂语法规则;可显示全文、图像和多媒体信息;采用客户机/服务器模式,可提供用户界面最大弹性;书目记录之间使用超链导航;用户可通过 Internet 检索远程数据库;具有搜索引擎相似的检索特性;能链接全文;可通过一个检索界面获取所有电子信息资源(如光盘、网络资源)等特点<sup>[6~8]</sup>。目前,WebPAC 正在发展之中。

### 1 WebPAC 书目记录内容的研究

在改善 WebPAC 书目记录内容方面,未来的 WebPAC 系统应能查询实际馆藏和虚拟馆藏,提供作品的知识内涵、章节和目次、书后索引供使用者参考<sup>[9,10]</sup>。

对图书馆有所有权的馆藏,目前很多图书馆基本上都在 WebPAC 中予以全部揭示,如加拿大国家

联合目录 AMICUS WEB<sup>[11]</sup>,内容包括 1300 多家加拿大图书馆的期刊、图书、报纸、地图、计算机文档、缩微片和各种特定格式作品(如盲文、电影、录像等)馆藏信息。

对图书馆仅有使用权的馆藏,目前国外著名大学图书馆均将电子期刊的 MARC 记录纳入 WebPAC 系统<sup>[12]</sup>,通过它揭示电子期刊的特征与信息,并在 WebPAC 书目记录中提供电子期刊链接点;BIBSIS 是挪威的大学图书馆、国家图书馆和研究图书馆的共享集成系统,目前已将馆藏数据库(即 WebPAC)与 ISI 的引文数据库、OCLC 的书目和全文数据库建立了链接关系<sup>[13]</sup>。

查询其他图书馆馆藏,建立虚拟联合目录,已不存在技术方面的障碍。所谓虚拟联合目录是指利用 Z39.50 协议连接各成员图书馆的 WebPAC 所形成的分布式联合目录。目前,不少发达国家都有虚拟联合目录相关计划,例如英国的 M25Link 计划,该计划从 1998 年起开始实施,目前已连接伦敦地区 125 所大学的 WebPAC 系统<sup>[14]</sup>;加拿大的 VCUC 计划,目前可通过统一检索界面单独或同时检索加拿大主要图书馆馆藏<sup>[15]</sup>;还有澳大利亚的 The Zedweb Project、德国的 DBV-OSI、意大利的 SIB YLIA、法国的 AQUARENE 和整个欧洲的 EUROPAGATE 等<sup>[16]</sup>。因此目前主要的问题仍在于研制成熟、精良的 Z39.50 协议的相关软件,克服虚拟联合目录的局限性,提高数据库的一致性和检索的准确性、系统的可

获得性等。

查询 Internet 资源方面,由于 WebPAC 所支持的原始数据库结构 MARC 在适应新的网络环境方面有其局限,因而目前基本上采用两种方案:

一是对 MARC 进行修订、补充和完善。如增设了 856 字段(电子资源定位与检索),该字段可包含 URL 以建立从书目记录到 Internet 的全文和多媒体资源的链接。但 URL 的不稳定,给目录维护带来严重问题。为此, IETF (Internet Engineering Task Force) 研制了 URNs (Uniform Resource Names, 统一资源名称) 标准,以赋予电子资源一个独立于信息资源存储位置的永久名称<sup>[17]</sup>; OCLC 研制了 PURL (Persistent Uniform Resource Locator, 永久统一资源定位器)<sup>[18]</sup>,以赋予电子资源一个固定的位置。目前由美国教育部主持的网络编目计划 (The Internet Cataloging Project) 已在其记录中使用 PURL,目的是创建一个全国范围的 USMARC 格式书目数据库,并评估其灵活性<sup>[19]</sup>。

二是选用其他格式代替 MARC 格式。SGML (标准通用标示语言) 允许标引和检索文献或文献的一部分,可支持布尔逻辑操作符、相关性排序,支持高级联机导航,允许根据需求在多个层面上使用等级关系处理信息,用户能控制标引、浏览和打印格式,对文献大小不施加任何限制。它独立于系统平台,可在彼此不相容的系统间交换信息而不遗失。已有越来越多机构在电子图书馆相关计划中选用其作为文件格式<sup>[20]</sup>。美国国会图书馆在 1995 年调查使用 SGML 标准编码 USMARC 格式的方便性后公布了 SGML 格式中定义 USMARC 数据的 MARC-CDTDs (Document Type Definitions),并在 1998 年研制出 USMARC 和 SGML 之间进行转换的软件<sup>[21]</sup>。目前已有采用 SGML 构建的实验型 OPAC 系统 (Cheshire )<sup>[22]</sup>。

## 2 WebPAC 书目记录显示和用户界面的研究

创建以更容易方式操作或显示信息是 WebPAC 的发展方向。Beheshti 设计的可浏览的、实验型 OPAC 系统——PACE (Public Access Catalogue Extension) 就是基于这样的思路构建的系统,它能将 MARC 记录表示成书籍的模拟影像,测试结果也表明 PACE 很容易使用<sup>[23]</sup>。

使用超链是 Web 的最显著特点,然而 Vrtiz-

Repiso 认为 WebPAC 的链接仅处于表面水平,只提供简单的通过共同特性(如 TI、AU)获取相关文件<sup>[24]</sup>。这是因为 MARC 是为描述和获得书目记录中所包含信息而设计的,而不是为同一记录字段之间以及单一文档不同记录之间建立链接而设计的。应采用其他格式代替 MARC 以更好地适应网络超文本结构,图书馆应在 OPAC 领域发挥更大作用而不能全部依赖商业性公司。

WebPAC 用户界面的设计应以用户需求为依据。目前每个 WebPAC 系统都有不同的特点和不同的显示,不同的显示可能包含相同类型数据,而表面上相似的显示可能包含不同的信息,这种混乱增加了用户利用 OPAC 的复杂性。Thomas 认为未来 OPAC 的设计主要依靠试验收集数据作精确地测试,并建议评估用户偏好,应给予更多机会使用户建立个性化的书目记录显示和用户界面<sup>[25]</sup>。

## 3 WebPAC 检索功能的研究

WWW 的普及,愈来愈多的图书馆提供 WebPAC 检索服务,WebPAC 的远端使用者将愈来愈多,而专业图书馆员不能给用户以解释、帮助,因此新一代的 WebPAC 系统应考虑通过下列途径提高检索功能:提供布尔逻辑检索;加强系统自动化侦测拼写错误;重视信息搜寻行为、加强联机指导;可将检索结果按相关程度进行排序;提供联机分类表和主题词表供用户浏览辅助用户检索;提供主题参照和联机索引以缩小或限制检索;提供整合的多种索引;支持模糊检索;具有自然语言检索能力;信息过滤;语音检索;对话式检索;智能检索代理 (Intelligent Search Agent),即用户将自己的信息需求交给智能检索代理,由它代替用户到各数据库中检索相关信息并返回给用户;相关反馈;应设计个性化 OPAC 系统,开展个性化服务,在用户查询时系统提供较符合用户兴趣和需求的信息<sup>[26,27]</sup>。其中有些功能(如布尔逻辑检索),很多 WebPAC 系统基本提供;而自然语言检索、相关排序等,很多 WebPAC 系统基本不提供。在整合上述检索功能方面有少数系统比较突出,如实验型 OPAC 系统——Oracle Libraries Multimedia OPAC,目前已具有高级提问处理、最佳匹配检索、相关排序等功能。其中高级提问处理主要通过自动合并检索词、截词、自动连接拼写检查、使用权重标引分析词等技术实现;最佳匹配检索主要通过计算提问词和文献中出现词的相似程度等技术实现<sup>[28]</sup>。

要进一步整合上述检索功能,目前主要采用两种途径:一是利用搜索引擎和专家系统来发展 WebPAC。搜索引擎发展很快,很多搜索引擎接受自然语言提问,具有模糊检索、相关排序和相关词扩展等功能。而专家系统具有启发性、透明性、高性能和灵活性等特点。Khoo 等利用搜索引擎和专家系统的特点建构的 WebPAC—E Referencer,结合搜索引擎诸多特性,使用 Z39.50 协议和专家系统界面,能处理用户自然语言提问;能将用户提问词和美国国会图书馆主题词匹配,并可自动扩展检索;能基于用户检索结果相关反馈来进一步限定检索和构建新的检索策略<sup>[29]</sup>。目前正在研制浏览检索界面,以允许用户通过浏览美国国会图书馆主题词和分类号进行检索。二是利用 SGML 或 XML 格式作为数据库的原始结构。Larson 等采用 SGML 构建的实验型 Cheshire OPAC 系统,能帮助用户构建有效检索提问,支持布尔逻辑检索、数据库浏览、相关排序和相关反馈等功能<sup>[30]</sup>。

#### 参考文献

- 1 Karen Markey. Subject searching in library catalogs: before and after the introduction of online catalogs. (Dublin, OH: OCLC, 1984)
- 2 Jamshid Beheshti. The evolving OPAC. *Cataloging & Classification Quarterly*. 1997, 24(1/2)
- 3 Charles R. Hildreth. Beyond Boolean: Designing the next generation of Online Catalogs. *Library Trends*. 1987, 35
- 4 叶乃静. 线上公用目录之使用研究:文献探讨. *大学图书馆*, 2000, 5(1)
- 5,6 Vinh The Lam. Online Catalogs—User Interface. URL: <http://www.leaf-vn.org/newsletter4.html>
- 7 陈惠瑜. 我国大学校院 WebPAC 介面与特性之评估比较. *大学图书馆*, 2000, 4(2)
- 8,16 曾元显. 架构在 WWW 与 Z39.50 上的近似自然语言 OPAC 检索系统. *大学图书馆*, 1998, 2(4)
- 9 黄倩如. 我国大学图书馆线上公用目录期刊功能与画面之研究. 硕士论文. 台湾大学图书馆资讯学研究所, 1998
- 10 林呈潢. 现代图书馆目录的功能与角色. *大学图书馆*, 1998, 2(2)
- 11 <http://www.amimus.nlc-bnc.ca/aaweb/amilagine.html>
- 12 邵晶. 对我国大学图书馆虚拟馆藏资源组织与揭示的思考. *大学图书馆学报*, 2001(1)
- 13 BIBSYS at the turn of the century the road to information. URL: <http://wgate.bibsys.no>
- 14 <http://www.m25lib.ac.uk>
- 15 <http://www.nlc-bnc.ca/816/index-e.html>
- 17 林海青. 数字图书馆的信息组织. *中国图书馆学报*, 2000(1)
- 18 Stuart Weibel, Erik Jul, Keith Shafer. PURLs: persistent uniform resource locators. URL: <http://www.oclc.org/oclc/purl/summary> (1996)
- 19 OCLC Online Computer Library Center, Inc. Building a catalog of Internet-accessible Materials. URL: <http://www.oclc.org/oclc/man/catproj/overview.htm> (1996)
- 20 Vrtiz-Repiso. Web-based OPACs: between tradition and innovation. *Information Technology and Libraries*. 1999, 18(2)
- 21 Library of Congress. Network Development and MARC Standards Offices, MARC-SGML Conversion Utilities. URL: <http://www.loc.gov/marc/marcsgml.html>
- 22,30 Ray R. Larson, Jerome McDonough, Paul O'Leary, Lucy Kuntz. Cheshire: designing a next-generation online catalog. *Journal of the American Society for Information Science*. 1996, 47(7)
- 23 Jamshid Beheshti. Public access catalog extension (PACE): a browsable graphical interface. *Information Technology and Libraries*. 1996, 15
- 24 Vinh The Lam. Online Catalogs—User Interface. URL: <http://www.leaf-vn.org/newsletter4.html>
- 25 David H. Thomas. The effect of interface design on item selection in an online catalog. *Library Resources & Technical Services*. 2001, 45(1)
- 26 Karen M. Drabenstott. Enhancing a new design for subject access to online catalogs. *Library Hi Tech*. 1996(4)
- 27 曾元显. 新一代资讯检索技术在图书馆 OPAC 系统的应用. *大学图书馆*, 1997(3)
- 28 Finbarr Joy. Extending the third generation OPAC. *ITS News. Journal of the Library Association Information Technology Group*. 1995, 32
- 29 Christopher S. G. Khoo. EReferencer: transforming boolean OPACs to Web search engines. URL: [http://www.ifla.org/ifa65/papers/010-143\\_e.htm](http://www.ifla.org/ifa65/papers/010-143_e.htm)
- 黄碧云 中南大学湘雅医学院情报学硕士, 讲师。通讯地址: 长沙市。邮编 410078。
- 方平 中南大学湘雅医学院信息管理系教授, 硕士研究生导师。通讯地址同上。

(来稿时间: 2002-01-04)