

基于知识库建设电子资源 集成系统

邵 燕

中国科学院文献情报中心 北京 100080

中国科学院研究生院 北京 100039

[摘要] 从4个方面分析图书馆目前的用户需求,同时结合 Serials Solutions、TDNet、1Cate 等国外电子资源集成系统的实际经验,提出基于知识库的电子资源集成系统框架。从三方面介绍其支撑平台——知识库的内容,并分析知识库在集成系统内的作用,在此基础上详细介绍系统的各功能模块,以期对实际建设起到一定的借鉴作用。

[关键词] 电子资源管理 跨库检索 集成 个性化 知识库 开放链接

[分类号] G250

Construction of the Integrated System for Electronic Resources Based on Knowledge Database

Shao Yan

Library of Chinese Academy of Sciences, Beijing 100080

Graduate School of Chinese Academy of Sciences, Beijing 100039

[Abstract] This paper analyzes the present users demand and illustrates the function of the integrated system for electronic resources by means of the digital resource systems such as Serials Solutions, TDNet, and 1Cate, etc. Based on this, the paper also presents the framework of an electronic resource integrated system and tries to go into details on its components, especially on the knowledge database which not only constitutes the foundation of the whole system but also forms an interface between resources and services.

[Keywords] ERM federated searching integration personalization knowledge database OpenURL

近年来,信息技术给图书馆带来了从未有过的冲击和影响,图书馆的资源和服务都发生着历史性的改变。目前,国内各高校图书馆及科研机构均呈现出数字资源大幅度增加、馆藏结构明显变化的趋势。以中国科学院文献情报中心为例,截止到2004年底,提供服务的电子期刊有10290种,大型全文数据库16个,此外还有一定数量的文摘和引文库。

丰富的资源在给用户的科研学习带来便利的同时,也给图书馆管理带来了极大的挑战——如何将形态复杂、种类繁多、分布广泛、平台相异的海量信息资源更好地揭示给用户,提供更优质的服务?目前,图书馆采取了一些措施来缓解上述矛盾,如通过馆藏OPAC系统对资源进行统一揭示,开发跨库检索软件,进行网络导航服务,提供OpenURL链接等。但是每种集成方式均有其局限性,如何将上述集成方式结合起来为用户提供更深层的集成化资源与一站式服务,是非常值得我们研究和思考的。

1 用户需求

要想建设一个真正服务于学习科研的信息环境,首先必须明确用户需要什么?用户需求是我们服务的根本所在^[1]。用户的需求大致可以归为4个方面:

1.1 方便地获取资源或相关书目信息

用户希望能够利用图书馆目录系统快速获得所需资源书目信息,同时还能便捷地访问相应的资源,如电子期刊数据库、网络资源等。有时用户可能并不清楚自己需要什么,图书馆应该能推荐一些相关的和适合的资源(比如相关主题链接等)给用户。

1.2 统一的检索界面

用户并不想去逐一熟悉各个数据库的界面,掌握其检索方法、使用技巧等,他们需要的是能够通过一次检索获取所有相关资源信息,并且是经过分析、去重后精练的结果,以便更快地发现所需资源。

1.3 资源与服务集成的平台

用户获取信息的途径一般有:查找相关信息——获取全文或相关资源——原文传递——馆际互借。对很多图书馆来说,这些步骤是割裂的,但用户需要的是通过一个入口,获取不同的电子资源和各种服务。比如图书馆没有电子版文献,系统会在用户发出原文传递请求后自动填写相应表单,提供相应的服务。

1.4 个性化界面

用户往往需要个性化服务,像能够定义自己常用的资源、保存检索式以备重复检索使用、获取研究领域的最新进展等。

明确了用户需求,也就明确了图书馆的建设方向——资源和服务集成化的信息系统。为了实现这个目标,我们还有必要对现有电子资源管理系统进行深入的分析研究,从而提出有价值的理论框架来指导实践。

2 国外电子资源管理系统

目前国外出现了很多功能比较强大的电子资源管理系统,在一定程度上满足了图书馆集成资源和服务的需要,因此对它们进行调研分析有助于我们完善自身的系统。

2.1 Serials Solutions^[2]

Serials Solutions 是为图书馆员服务的公司,该公司着力于解决电子资源使用和管理的相关问题,为图书馆员和用户提供更多的便利。其产品主要有 E - Journal A. M. S.、Full MARC Records、Article Linker、Central Search、Data Management Services 5 个部分。

E - Journal A. M. S. 即 E - Journal Access & Management Suite(电子期刊使用和管理套装),包括电子期刊门户、A - to - Z 期刊列表、重复分析、期刊链接、主题浏览、题名检索、乌利希期刊指南网上链接、数据文档等功能。这一模块主要用于揭示电子期刊。

Full MARC Records Service 体现了 Serials Solutions 由图书馆员开发的优点,他们花了 4 年时间跟踪全球 100 多个数据提供商的 800 个数据库,参考了 CONSER 等的编目记录,并进行了格式规范,从而能够提供海量的、规范的 MARC 格式书目。

Article LinkerTM 是一个支持开放链接的解析器。除了具有一般开放链接解析器的功能(多种类型资源之间的链接,提供扩展服务,整合电子馆藏和印本馆藏,提供最合适的链接方式等)外,还有对检索结果进行去重、提供相关主题链接等特色功能。

Central Search 集成检索功能强大,并对检索结果进行合并、去重操作;用户可以定制个人页面;通过 Z39.50、XML、HTTP 等协议支持,可同时检索多种资源,如全文、文摘和免

费数据库及其他图书馆 OPAC 等;支持多种授权模式,即具备多种代理机制,符合多种认证要求。

2.2 TDNet^[3]

TDNet 是 Teldan 公司开发的电子资源解决方案(E - Content Solutions)的组成部分,也是服务于图书馆等信息机构的电子资源管理系统。其针对电子资源的管理主要提供两种解决途径:Searcher & Analyzer 和 Journal Manager。此外,还提供了一些电子期刊工具,如 TOURResolver、Catalogue Updater 等。

eContent Searcher & Analyzer 是 TDNet 提供的集成检索系统。该系统的特色体现在检索结果的处理上,基于已申请了专利的先进数据算法,该软件具有三个独特的功能:根据检索结果中每条记录的首页信息进行结果的自动分类,自动文摘,自动提示“相关主题”资源。

Journal Manager 主要是进行资源管理和现期通告服务。系统拥有 65 000 多种期刊,35 000 多条目次记录,500 多个数据库。系统每周更新,能集成图书馆 OPAC,支持文献传递和详细的馆内使用统计功能。此外,用户可定制“我的 TDNet”,创建个人档案以获取最新资讯通告等。

TOUR - TDNet OpenURL Resolver 是支持开放链接协议的上下文敏感链接系统,系统不仅能识别支持开放链接的资源,还能够对其他非标准化、非结构化的内外部资源进行操作,同时与 eSearcher & Analyzer 进行集成。

2.3 1Cate^[4]

1Cate 是由 Openly Informatics 公司开发的,它为信息领域开发链接服务器软件提供了一个开放的、支持互操作的平台。1Cate 是 1 Click Access To Everything 的缩写,即一次点击获取所需全部信息。系统包括字母顺序期刊列表、Journal-Seek 链接数据库、CrossRef 链接模块、1Cate 链接服务器。

值得一提的是开放链接知识库。该库存储着来自 1 200 多家出版社的 75 000 多种电子期刊的描述信息和链接信息。此外,系统还包括了 600 多个款目表(不同数据库的期刊款目),以通用款目格式(UHF, Universal Holdings Format)存放,增强了通用性,内容基本上涵盖了目前顶尖的电子期刊集成商(Blackwell - Synergy、EBSCO Host 等)的产品。

此外,笔者还对 Ex Libris 公司的电子资源管理软件系列——SIRSI 系统、EBSCO 公司和 SWETS 公司的相关系统进行了调研,限于篇幅在此不一一列举。

综合分析之后,笔者发现上述系统可谓各有千秋。比如 Serials Solutions 的 MARC 编目数据比较权威,并且覆盖范围很广,但是其统计功能和个性化服务比较薄弱;TDNet 系统的集成检索功能强大,且提供定制服务;1Cate 系统主要提供开放链接相关软件,但其知识库数据丰富,且进行了规范处理,通用性很好;Ex Libris 公司的 SFX 是辅助图书馆实现开放链接功能的软件;MetaLib 是一个集成检索系统,能进行身

份认证,还可与 SFX 集成;DigiTool 是针对电子资源的编目和元数据管理而开发的;Verde 则是要解决图书馆各个工作流程(评价、选择、购买等)中电子资源的相关问题^[5]。

3 基于知识库的系统建设

综合上述各系统的优势,我们设计出了基于知识库的电子资源集成系统的整体框架(见图1),下面对其中一些功能模块进行深入分析,以期对图书馆完善自身系统起到一定的借鉴作用。

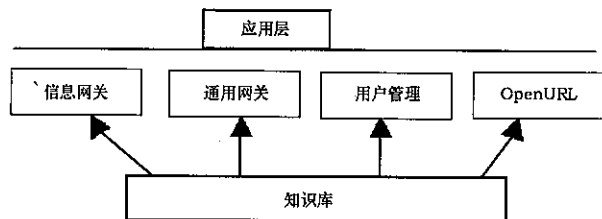


图1 电子资源集成系统整体框架

知识库是底层数据平台,主要有信息网关、通用网关、用户管理和开放链接4个功能模块。应用层是提供给用户的服务界面。

3.1 知识库的内容

3.1.1 描述性信息 主要是关于各个资源的信息,比如:资源名称、主题范围、存放位置、出版商等描述性信息。I Cate 系统的知识库能够深入到每期和每篇的层次上对资源进行描述,比如:刊名、ISSN 链接到期刊中单篇文献的说明、分类体系等。它们是知识库内区分资源的标识,同时系统也能根据信息对资源进行分类,在不同的上下文环境中以不同的方式给用户提供服务。

3.1.2 功能性信息 主要是关于如何获取资源的信息,比如资源的访问方法、资源的物理形态、逻辑形式、所用的字符集等。有了这些信息,系统的通用网关就可以生成符合资源说明的检索式,发送这些检索式给相应的系统,然后对结果进行处理。

3.1.3 上下文敏感链接信息^[7] 上下文敏感链接信息是包括以下4方面内容:Source、Source Services、Targets 和 Object。

Source 是一个可以创建 OpenURL 的信息源,也是用户获取资源的入口或门户。通常指摘要或引文数据库,例如 SilverPlatter 的 Medline 数据库、ISI 的 Web of Science 等,也可以是全文数据库、电子出版文档、学术论文索引库等,还可以是图书馆的馆藏目录系统。

Source Services 是一个可以提供服务的信息源,比如:查询本地馆藏、获取全文、文献传递、查看相关图书评论等。

Targets 是带有明确链接语法的网络资源,包括:全文期刊、OPAC、文献传递表格、搜索引擎和网上书店等,一般是用户想要查询的目标信息源。

Object 是指用户想获得的对象,比如期刊、论文等。不同的信息获取权限不同,每次发送链接请求时都要按照 OpenURL 协议的规则,生成合适的链接,避免将图书馆没有权限的资源提供给用户。

3.2 系统功能模块

3.2.1 信息网关 其实就是资源目录库,各种资源的核心目录,相当于各种资源的导航系统。图书馆可以配置核心目录,包括:资源名称、存放地点(URL)、主题范畴、使用的语言等。针对电子资源的不同特点,参考 CONSER 编目规则进行 MARC 著录,也可以考虑使用 DC 元数据格式著录资源。提供资源推介服务,将用户感兴趣的资源推送到用户的个人目录中去。

3.2.2 通用网关 实现跨库检索,并对检索结果进行相应的处理和提炼,主要包括5个功能。

- 处理用户提交的检索式,并获取来自不同数据库的检索结果。将查询语句转换为目标资源要求的格式:遵循 Z39.50 协议的系统,通过规范的 Z39.50 接口实现跨库检索;对于不规范的系统,依据其协议的公开程度分类处理;对于一些无法确定协议类型、通过网络服务的系统,则采用基于 HTTP 协议的“页面分析技术”实现跨库检索^[8]。

- 对检索结果集进行格式一致性处理,在调整和转换之后呈现给用户。

- 按照用户的需要进行检索结果的再次集中,其类目规则和排序方法由用户来确定。

- 查重。在所得检索结果集中识别重复的记录,生成去重表。

- 自动分类。可参考 TDNet 系统的功能,通过语义分析,推断某一文件中的相关关键词,并将检索结果自动归为几个概念簇,便于用户更快地找到最相关的条目。

3.2.3 用户管理 包括两个层面:通过机构内的用户库(知识库的一部分)认证用户身份;在集成商系统中,图书馆作为数据库购买方的权限。

机构成员和临时访问者拥有不同的使用权限,不同的机构成员权限级别也不相同,系统根据不同的组级对用户开放不同许可权的资源和服务。同时,机构成员可以定制个人数据库、个性化检索环境,提供将检索结果保存在电子书架上(E-shelf)、定制通告等服务。

当用户使用机构权限外的数据库时,涉及到机构在数据库集成商处获得的权限。如果用户拥有个人的账号,那么系统可以将存在知识库中的用户账号密码信息调出,与远程数据库系统交互;如果用户是在自己图书馆以外的地方进行检索,而其所需数据库是通过 IP 进行权限控制的,那么系统可以设置代理服务将用户重新定向到图书馆的 IP 地址范围内,从而获取相应的资源授权。

3.2.4 OpenURL 解析服务器 OpenURL 是一种协议,规定

了信息服务提供者之间传递对象的元数据格式,形成一种开放的、上下文相关的链接框架结构。该协议允许信息源(如 Elsevier 或 ProQuest) 传送特定对象的元数据或者是元数据对象的相关信息(如期刊 ISSN 号、标题、作者、期卷、刊号、页码等)到某机构的网络服务组件上(通常是一个服务器,如 SFX 服务器)^[9]。

解析服务器针对某个文献提供获取全文的最佳链接,以及为用户提供上下文敏感的链接服务,这些链接服务包括全文链接、提示印刷文献的具体馆藏信息、导航到其他数据库(如引文数据库)的相关信息、创建馆际互借或文献传送请求等^[10]。

3.3 知识库与系统其他模块的关系

之所以将知识库从系统的功能模块中独立出来,是为了便于维护、增强系统的可扩展性和开放性。系统设计不可能一劳永逸,因为资源的种类和数量都是在不断的变化中,用户的需求也随着技术的发展而改变着,这就造成了资源和资源配置信息的变化。这些信息如果和资源服务绑定,我们进行维护的工作量就非常大,并且可能要更改系统结构。

因而,从一定意义上来说,在整个系统中知识库提供了资源与服务对外的通用接口,通过它,资源和服务发生变化后,只需要对其配置信息进行修改,而系统其他部件不需要做大的改动。这样既保持了原有系统结构的稳定性,又便于系统扩展新的功能和服务。

4 结 语

[作者简介] 邵 燕,女,1981年生,硕士研究生,发表论文2篇。

(上接第40页)

高,网站因此为公司带来了可观的经济效益。

2.4.2 馆藏媒体与商业网站间的互动 为了更好地利用馆藏的传统媒体,图书馆可以将其与商业网站进行互动合作,互惠互利。一方面可以使馆藏媒体发挥更好的传播效果和经济效益;另一方面又可以使商业网站获得其缺乏的信息资源。例如,1999年9月,新华社与新浪网合作设立庆祝国庆专题网站,由新华社提供信息资料,新浪网进行技术设计。该网站一时成为最热门的专题网站,取得了极佳的效益。又如,联想集团一开始建网站时就与全国数百家新闻单位签定了合作关系,利用传统媒体的信息优势来提升自身实力。

2.4.3 自建网站与商业网站间的互动 拥有政策优势和资源优势的图书馆自建网站与自主经营和运作的商业网站进

在对国外若干实际系统进行综合分析并在对相关技术深入了解的基础上,提出了基于知识库的电子资源集成系统框架,希望能够对图书馆进行系统建设起到一些借鉴作用。当然,系统框架仍然存在一些有待补充、完善的地方,比如:联盟图书馆和多分馆系统的特殊问题;资源使用统计问题;各功能模块的数据间如何更好地相互对话,实现最小重复等,这些都有待于更进一步的探讨研究。

参考文献:

- [2004-11-20]. http://www.oclc.org/research/staff/dempsey/recombinant_library/
- [2005-03-17]. <http://www.serialsolutions.com/products.asp>
- [2005-03-17]. <http://www.tdnet.com/site/page.asp?ID=449A&parent=449>
- [2005-03-17]. <http://www.openly.com/1cate/>
- [2005-03-17]. <http://www.exlibrisgroup.com/>
- [2004-10-15]. <http://www.openly.com/journalseek/>
- 李富玲,卢振波. SFX——信息资源整合新工具. 现代图书情报技术,2002(6):70-71
- 姜爱蓉,王平,郑小惠. 分布异构资源整合管理系统的技术特点和应用趋势. 现代图书情报技术,2004(1):3-4
- 沈艺. Open URL 及其应用. 现代图书情报技术,2004(1):30-32
- [2005-03-12]. <http://www.exlibris.co.il/metalib.htm>

行互动合作,既可以克服图书馆自建网站在经营和运作上很难做到快速应变和灵活决策的弊端,又能丰富商业网站的信息资源,一举两得,互惠互利,有利于双方的发展。例如,浙江在线通过与新浪的合作,就扩大了其在全国的影响,而新浪也得到了浙江在线的信息资源。

参考文献:

- 吴杉. 网络时代迅速崛起的网络媒体. 焦作大学学报,2004(1):108-110
- 汪苑菁. 网络媒体优劣谈. 新闻世界,2004(2):37
- 闵大洪. 中国社会变革与新媒体使用. Internet 信息世界,2001(12):78-82
- 孙琳娜. 网络时代的媒体变革与思考. Internet 信息世界,2001(12):85-88

[作者简介] 邹忠民,男,1946年生,研究馆员,发表论文60余篇,出版著作4部,参编著作1部。

丁元江,男,1960年生,副研究馆员,发表论文23篇。