

高校图书馆资源整合的技术分析*

罗明英, 岳付强, 坤燕昌

(西昌学院, 四川 西昌 615013)

【摘要】本文对目前高校图书馆普遍存在的几种资源整合模式进行技术分析,阐述了各整合技术的优缺点。为设计和实现下一代智能系统整合,规范元数据的收割,建立数据仓库,进行数据挖掘提供参考。

【关键词】图书馆;元数据;资源整合模式;智能系统

【中图分类号】G250 **【文献标识码】**A **【文章编号】**1673-1891(2011)04-0083-03

网络技术、信息数字化技术的发展,改变了信息的获取、拥有、利用处理、交流的方式。图书馆不仅成为重要信息的生产基地,而且也成为多种信息资源的集散地,是信息发现、搜索、捕捉的一个重要导航站点。图书馆不仅仅是书的馆藏地,而且还要提供不同载体、不同类型的数字化信息资源,一方面给读者利用信息资源提供方便、快速的途径,另一方面,这些多样性资源的采购、管理、使用也提上了图书馆的工作日程,只有充分了解数字化资源的技术构成,才能整合不同数字出版商提供的资源成为图书馆的信息,节约馆内成本,取得较高的投资回报。

1 高校图书馆资源整合的诠释

高校图书馆为了保障学校教学、科研的需要,每年都在提高电子资源的建设力度,从存储方式上,电子资源有本地镜像和远程序包库。从类型上,有期刊、电子书、视频。从操作上,有不同的阅读器。数字资源整合,是指依据学校的实际需求和需要,把不同来源和通信协议的资源链接上,使各种类型、各种格式、使用频繁的数字资源本地化,进入数据仓库,实现无缝链接。通过数字资源整合后的系统,具有集成检索功能,是一种跨数据库、跨文档格式、跨平台的系统。

2 高校图书馆资源整合的背景

随着计算机技术、网络技术、通讯技术的广泛应用,无纸化办公已成现实,数字出版业快速发展。各高校数字资源建设加快进程,一批中小数字化商纷纷推出各有特色的数字化产品,如期刊CNKI、万方数据、维普;图书的超星、APABI、书生;视频的网上报告厅、软件通、银符考试。

高校资料配置往往根据各高校的定位,使用经费来决定,211、985高校不仅购买国内的数据库,还有国外的专业数据库,一般的本科院校大都只购买国内的数字资料,职业技术学院近年来也开始购置

一些国内常用的数据库。在资源整合方面,已有“超星”推出的“读秀”系统,四川高校大多数使用这个系统进行资源整合。

面对图书馆信息资源现在分散无序的状态时,读者最想要的是什么呢?其实读者真正需要的就是想有一个类似Google的搜索框,一个回车就能得到图书馆拥有的所有相关数据。所以现在就要求图书馆需要应用资源整合的观念,形成可以整合不同网络资源数据库的统一用户界面,从而提高数字资源的利用率和检索率。

3 现行的几种资源整合系统

3.1 资源整合系统OPAC的技术分析

OPAC的全称是Onling Public Access Catalogue,专业称谓是公共电子书目,构建在OPAC系统下的馆内信息资源整合是现在图书馆最基本的整合方式,它需要整合的内容主要包括馆藏书目、电子图书、电子期刊等,实现馆藏书目信息资源与数字资源的一体化整合。

OPAC系统实现方式是将数字资源进行MARC编目。这就要求数据商将他们的数字资源的对应的MARC数据提供给图书馆,在所提供的MARC数据里面最重要的是856字段(电子资源地址与存取字段),用来记录数字资源的存取方式和存取地址。只要MARC数据中包含了856字段,那么当导入到OPAC系统时在OPAC系统中的856字段就会有相应的数字资源的链接。当用户在检索到目录信息后,即可通过相应的链接获得书的全文信息,获取与之相关的期刊、图书以及学位论文等各种资源。

此外,OPAC将通过Z39.50协议引入hook机制,根据SFX来链接后台资源服务器,以提高本地馆藏资源扩展服务,实现不同信息源之间的连接,并与其它图书馆的OPAC实现异构平台的互检,OPAC资源的互相操作,以达到所有图书馆资源的完全整合,如图1所示。

收稿日期:2011-06-15

*基金项目:本文的研究成果来自西昌学院“高校图书馆资源整合系统的研究”(项目编号:ZZSSA0732)。

作者简介:罗明英(1964—),女,四川冕宁人,副教授,博士,研究方向:神经网络、信息库数据库与数据挖掘、数字图书馆。



图1 资源整合后的OPAC系统的用户界面

资源整合系统 OPAC 具有功能齐全、操作简便、术语简明的特点,是用户查找和利用图书馆资源的一个很好的途径。但是由于各数据商的数字资源的 MARC 标准并不完全统一,甚至有的数据商就没有对应的 MARC 数据,所以这种整合模式会降低整合的准确率和准全率。

3.2 导航系统的技术分析

资源导航系统是对图书馆现有的各种数字资源的 URL 建立导航数据库,该平台提供多种入口方式,例如按字母、按文献类型、按学科类别等方式来进行检索。该导航软件系统是辅助用户全面了解数字资源,提供给用户浏览或按某种特征来进行检索入口。为了使该导航系统能达到预计的功能,要明确其揭示的内容,该内容的详细程度决定了资源导航系统具备的功能。资源导航系统通常都具备分类导航功能、字顺导航功能、关键词检索功能三个基本功能,利用这些功能辅助用户迅速定位到其需要的数字资源,并提供检索入口。

资源导航系统主要是引入了分类组织法,成功屏蔽了由于各种数据库资源本身差异带来的不便。同时利用结构化的资源列表来方便用户快速定位到所需的相关主题处,特别适合那些对所需内容并没有明确概念的用户。但是严格的来说这种整合只是将图书馆庞杂的信息资源进行了排序整合,还算不上真正意义上的资源整合,读者仍需要通过导航系统分别访问每个数字资源系统,如图 2 所示。



图2 导航系统资源整合后的用户界面

3.3 异构检索平台的技术分析

跨库检索是现在高校图书馆信息检索的重要构成部分,它主要用来实现分布式异构数据库的检索。跨库检索常常也称作联邦检索(federated searching)、集成检索(integrated access)、多数据库检索(multi-database searching)、分布式检索(distributed searching)、一站式检索(one-stop searching)等,它在统一的检索环境中实现对不同数据源和不同数据库的查询,并对检索结果去重和排序等处理,在一个完全统一界面中显示检索结果。其优点是用户可以避免切换各类检索工具,不必关心各种检索工具使用的不同检索语言以及检索途径,准确、全面、快速地查询众多数据库,并可对检索到的资源进行合理的管理。这种整合模式是目前主流的整合方式,它大体上又分为两种形式:元搜索/技术型和仓储/聚合型。

3.3.1 依赖跨库检索工具的元搜索/技术型

这种异构检索平台的整合主要是通过检索代理工具实时对分布异构资源进行整合来实现检索,其主要的实现方式有:(1)利用相关协议主要有 Z39.50 和 OAI 协议。两种协议都可以实现异构数据库检索,但目前大部分网络数据库都没有专门提供针对两种协议的检索接口,应用仍有一定局限性。(2)HTTP 页面分析,通过对 URL 和 Web 页面的分析和编程实现,这种方式是目前比较常用的方式。它的实现原理是首先整合平台对各数据商的数据库页面进行自分析,找出其检索的入口并将其入口写入自己的数据库中,这就相当于在整合平台中建立了一个中间层。当得到用户检索请求后,交给一个叫 Agent 的服务器端程序,Agent 程序将根据不同的数据库,把用户请求转化为符合要求的格式,再将检索请求发送到相应的数据库。在得到数据库的应答后,将来自不同数据库的检索结果转化为统一的格式,并将其发送到客户端浏览器,最终把检索结果显示给用户。目前在国内主要有同方 USP、TRS 等整合平台,国外主要有 METALIB(如图 3 所示)、SFX 等整合平台。



图3 Metalib 检索界面

3.3.2 依赖元数据收集的仓储/聚合型

依赖元数据收集的仓储/聚合模型是目前应用较多的跨平台系统之一。该模型主要通过通过对多个全文库数据库,按照一定标准进行信息标引后,组成一个元数据集,其通常是用数据仓库的方式储存。通过一个Web服务器的发布系统与客户端进行交互。用户将看到一个集成后的多源数据查询系统,得到一个对原始数据进行描述后的全文库信息。通过特定连接,用户将直接从原始数据源中得到所需用的原文数据。简单的说,就是优先收割/收集异构资源的元数据,进行聚合存储到元数据仓储后,提供统一的检索和服务。

3.3.3 两种异构检索平台模型的比较分析

元搜索/技术型跨数据库整合模式优点:即时性比较好。劣势:检索效率比较低,用户需要等待较

长时间;无法对检索结果完全的过滤和排序,检索结果界面不一、数据重复;相当数量的资源提供商不遵循标准协议,有些数据库的检索语句经过加密处理,因此不可能用中间层的方式整合所有的资源。

元数据收集的仓储/聚合型跨数据库整合模式优点:数据经过收集转换后不仅格式统一,而且结构清晰,检索结果界面统一,无重复,排序效果好,检索效率高速度快。劣势:元数据收集整理技术难度高、工作量大,需要大量投入人力物力。

4 结论

本文主要对目前高校图书馆普遍存在的几种资源整合模式进行了技术分析,阐述了各整合技术的优缺点。主要为设计和实现下一代智能系统整合,规范元数据的收割,建立数据仓库,进行数据挖掘提供参考。

注释及参考文献:

- [1]王瑜.图书馆信息资源整合与应用[J].辽宁工程技术大学学报(社会科学版),2007(1):66-68.
- [2]赵凤丽.我国图书馆OPAC功能及其资源整合现状分析[J].图书馆界,2006(4):39-42.
- [3]傅永阳.图书馆信息资源整合的不同模式与技术分析[J].情报资料工作,2006(5):58-60.
- [4]李晓莹.图书馆异构数据库检索系统功能分析[J].情报杂志,2007(2):27-29.
- [5]廖立顺.浅谈高校图书馆数字资源整合[J].龙岩学院学报,2006(12):119-121.
- [6]毛垣生.异构跨库检索技术理论和实践[J].情报学报,2006(10):313-316.
- [7]徐春玲.基于Agent的异构数据库统一检索系统设计与实现[J].长沙职业技术学院学报,2006(2):107-109.
- [8]叶华,张森.一种异构数据库间的XML接口方案[J].计算机应用,2001(6):10.
- [9]ExLibris LTD.OpenURL Overview.http://www.exlib-risgroup.com/sfx.Htm.2007-12-01.
- [10]图书馆数字信息资源整合技术研究进展[J].图书馆情报,2008(7).
- [11]Carl Lagoze.The Open Archives Initiative Protocol for Meta-data Harvesting[EB/OL].http://www.openarchives.org.2007-12-01.
- [12]CrossRef.org[EB/OL].http://www.CrossRef.org.2007-12-01.

Technical Analysis of Resources Integration of the University Library

LUO Ming-ying, YUE Fu-qiang, KUN Yan-chang

(Xichang College, Xichang, Sichuan 615013)

Abstract: This paper does technical analyses on several common resources integration of university library, and expounds the advantages and disadvantages of each integration technology. It provides the reference for the design and realization of the next generation of intelligent system integration, regulating reaps of metadata; establish the data warehouse and data mining.

Key words: Library; Metadata; A model of resource integration; Intelligence system