

非书资源管理与图书馆 数字化管理体系的构建

叶 岚

(苏州科技学院, 江苏 苏州 215011)

摘 要: 资源整合是数字图书馆建设的重要基础工程。理想的数字信息资源整合是把各种信息资源透明、无间隙地联在一起, 让使用者感觉如同只在一个资源系统中操作。如何将现有的图书馆联机公共检索目录(OPAC)与各种数字资源、电子资源及网络资源合理的整合在一起, 实现资源的有效利用, 已经成为现代图书馆管理的焦点。而经过整合的资源可以充分发挥计算机和网络的强大功能, 使信息检索效率得到快速提高。通过从管理系统选择、硬件配置和管理、资源加工与处理等几个方面的介绍, 以苏州科技学院图书馆非书资源与服务平台为例, 提出了非书资源在 OPAC 中存在的一些问题, 提出了解决问题的相关方法。

关键词: 非书资源; 整合; OPAC; 图书馆

中图分类号: C939

文献标志码: A

文章编号: 1009-1971(2009)03-0099-05

一、信息资源整合的起源及发展

20 世纪 80 年代, 信息资源整合主要是针对异构数据库的集成来进行探索的。主要方法是, 采用紧密耦合的方法对异构数据库进行集成。由于这种方法采用集中控制分布数据库, 每次局部系统的改变都会引起对全局系统的修改和重新构造, 使得系统不但开发成本太高, 而且后期维护困难。以后开始逐渐采用松散耦合方法, 出现了联邦式数据库集成框架和多数据库语言方法集成架构, 每个数据库提供一个输出模式供其他数据库进行共享。

20 世纪 90 年代初期, 随着网络技术的发展, 伴随着各种数据库访问标准和数据库互联产品的出现, 又相继涌现出很多数据库联合模型, 可为多数据库提供统一查询接口, 但多数据库语言集成会增加用户端负担。到了 20 世纪 90 年代中后期, 人工智能研究者也介入到信息资源整合这一研究领域。他们主要针对数据库存在语义异构问题, 提出基于语义层面整合的信息资源整合方法, 信息资源整合的范围也从数据库扩大到互联网资源。

我国的信息服务机构于 20 世纪 90 年代末开始了信息资源的整合实践, 主要着手对购进的大型数据库提供一站式检索服务, 即实现跨库检索。2000 年以来, 现实需求和技术变革驱动着数字资源整合系统的快速发展, 国内图书馆和软件商也开始关注和开发数字资源整合系统, 目前已研发出产品的有清华同方公司的 Union Search Platform、北京拓尔思公司的 Union Search-System 等。这些系统各有特色, 在体系结构、技术思路和功能实现方面也不尽相同, 如中科院文献情报中心自主开发了“CSDL 跨库集成检索系统”, 北京大学图书馆采用“CALIS 资源统一检索系统”, 南京大学图书馆采用的“一站式文献检索系统”, 上海交大采用的“跨数据库检索系统”, 华南理工大学和首都师范大学等图书馆选用了“清华同方异构数据库统一检索平台”。

二、图书馆信息资源的优化整合

首先是整合原则。一是科学性。信息是一种网状结构, 由众多结点和结点间的联系组成。结点是组成信息的最细成分; 信息关联是若干个信息因子间的联系。信息资源的优化整合针对

收稿日期: 2008-06-14

作者简介: 叶岚(1962-), 女, 上海人, 馆员, 从事图书馆数字化研究。

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. <http://www.cnki.net>

的是信息因子的有序化和信息关联的网络化。二是系统性。网络环境下图书馆信息资源由实际馆藏和虚拟馆藏两部分组成。只有系统地、连续地从馆内与馆外、国内与国外收集和积累各种数字信息资源,进行优化整合,才能不断地充实和发展图书馆的实际馆藏和虚拟馆藏。三是标准化。传统图书馆在信息资源建设过程中采用的软件差异很大,各馆的数据库建设也各自为政,其标准和格式不一致,无法在网上共享。因此,信息资源整合必须按标准进行。四是共享性。只有通过图书馆间的协同发展,形成互为补充、利用、推动的文献信息资源保障体系,才能提供网上信息服务,发挥信息网络的作用。五是效益性。社会效益指数字图书馆运行所产生的有益于社会进步的效果,给读者所带来的满意程度,对社会发展所起促进作用等。经济效益主要指图书馆对科技进步、宏观决策以及相关产业发展所起的作用。五是特色性。从网络整体出发,进行资源的合理配置,把图书馆网络建设纳入整个地区、国家和全球信息网络中,加强特色数字信息资源开发,建立特色数据库。七是安全性。采取必要措施,保证图书馆网络与信息资源的安全与正常运行。保护措施有:防火墙技术、VPN(虚拟专用网)技术、加密技术、网络病毒防治技术、跟踪检测技术等。^①

再者是整合目标。现代图书馆更加注重信息资源开发利用程度,最大限度地满足不同用户的需求。信息资源体系建设以网络为依托,将不同的馆藏资源共享,满足更大空间范围用户的信息需求。

还有整合中存在的问题。图书馆自动化系统发展到今天,虽然已经趋于成熟,但是 OPAC 公共检索功能仍有缺陷,而这种缺陷在目前的图书馆自动化系统里很难弥补(因为早期的 MARC 标准没有预测到图书馆的电子资源会发展到今天这样的程度)。目前的 OPAC 系统主要存在以下的缺陷:随书和单独发行的多媒体资源(光盘与磁带等)无法在 OPAC 直接使用;与书刊相关的电子资源(电子图书、电子期刊等)无法在 OPAC 查询到并直接使用。为此,本文结合实例

构建非书资源管理平台,对其中的技术问题提出了相应的解决措施。

三、非书资源管理平台操作系统的选用及技术指标性能

以苏州科技学院图书馆为例,选择了“畅想之星”非书资源管理平台作为图书馆非书资源的管理平台。这种资源的管理突破了传统的关系数据库中以记录形式来管理的方法,而采用资源对象的方法来管理,资源对象可以随时增加自己的属性(元数据),每种属性具有自己的方法去访问相应的物理数据(对象数据),引入了“主题对象”的概念,实现了基于内容的跨媒体检索功能,从而能够轻易实现各种非书资源的整合。

系统的整体框架采用基于 NET 的 B/S 架构,由于系统管理的对象数据量大(资料个数和占用的存储空间都很大)、数据类型各种各样(有音频 WMA、MP3 RM、交互式外语学习,视频 WMV、RM、AVI 电子图书 PDF、HTML、CHM、EXE、DOC 等)和处理过程复杂(自动生成压缩包、压缩包的在线浏览与动态释放、生成虚拟光盘 ISO 文件、底层驱动以及虚拟光驱的远程插盘等)的特点,使用普通的 B/S 技术无法满足要求,系统对 NET 和浏览器都进行了深入的编程(重新改写了 windows 系统的多个底层的类库),使该系统高效稳定的运行。^②

四、非书资源与馆藏目录的联系

目前大多数图书馆都有各种各样的数据化资源,如书目数据库、全文数据库、光盘数据库等,这些不同的数字资源系统有着不同的编码结构、表达方式及访问接口,导致其描述和组织标准的差异及检索途径和方法的不同;不同的数据库使用不同的检索软件,使得这些数据库系统的检索接口也风格各异,具有差异性和复杂性。用

①王红岭.网络环境与图书馆信息资源整合利用研究. <http://www.bjpopass.gov.cn/bjpopassweb/n9196c59.aspx>

②“畅想之星”非书资源管理平台使用说明书。

户在利用这些数据库时,需要在不同的网络环境之间穿梭漫游和在不同的信息空间来回切换,还需要掌握不同的检索软件的使用方法,这给用户利用数字资源带来很大的不便。从某种意义上讲,数字资源量越大,给用户造成的负担也就越重。因此,对馆藏数字资源进行有效的整合,以统一的访问接口及检索界面将信息提供给读者,使读者能够有效地利用图书馆信息资源,已成为新时期图书馆的一个重点工作。

OPAC (Online Public Access Catalogue) 即联机公共检索目录,是一种通过因特网对馆藏资源时行检索的工具,具有功能齐全、操作简便、术语简明等特点,也是历来读者查找和利用图书馆书刊资源的主要途径。OPAC 服务的全部数据都来源于馆藏记录,内容涵盖印刷型和数字型资源^[1]。基于 OPAC 系统的功能和地位,以现有馆藏资源为核心建设集各种电子资源为一体的 OPAC 系统具有一定的可行性,其结果也将为读者访问图书馆的丰富资源打开方便之门。因此,以 OPAC 系统为基础,整合其他信息资源,可以极大地提高图书馆现在所有信息资源的利用率。这种基于传统书目管理的整合模式,其技术比较成熟,国内很多图书馆已经实现了这种模式的整合。

馆藏目录是揭示、识别、检索馆藏文献的重要工具^[2]。在数字化时代,馆藏目录的建设面临着两个方面的任务:一是图书馆引进或自建了大量的数字资源,如何有效地将这些资源在馆藏目录中揭示出来;二是如何通过 Z39.50 协议集成不同图书馆的 OPAC 数据库。

五、非书资源在 OPAC 应用中的基本方法

资源整合本身就是一种服务,同时是构建各种服务的基础^[3]。

1 随书光盘资源整合

随书光盘服务系统是一个提供随书光盘网上下载、使用的服务系统。可以在 OPAC MARC 记录 215 字段@ e“附件”子字段中标注或揭示随书光盘信息,同时在 300 字段中加以描述和说明。当读者通过题名、ISBN、索书号等途径在 OPAC 系统中检索到某一本书时,OPAC 则以

215 字段的@ e 为输出内容,告知用户该书有无随书光盘。如果存在,则通过调用随书光盘服务系统接口,用户可以直接点击下载该书的随书光盘,从而实现 OPAC 系统与随书服务系统的整合。

为了对随书光盘进行科学合理的管理,建立的数据库应体现本馆馆藏信息,能方便地实现数据的录入、修改、查询、统计、打印、数据备份等功能;有基本的检索字段,如书名、作者、ISBN、索书号、关键字等^[4];能与图书馆自动化管理系统相结合,使网上用户不仅可以通过随书光盘数据库查询到所需光盘,还可以通过图书馆馆藏查询直接找到该书光盘,并能对光盘内容进行在线浏览或下载;用户界面要求交互性好,操作方便;能通过对用户 IP 地址的限制,实现对光盘使用范围的管理。目前市场上已经有这种比较成熟的管理软件,售价也不高,我们完全可以利用现成的软件作为开发基础,这样不仅可以减少不必要的人力物力浪费,还可以大大提高我们的开发效率。

2 基于 OPAC 系统的电子资源整合

图书馆用 MARC 格式来揭示图书信息资源,并通过 OPAC 系统实现文献资源的检索。随着电子资源的出现和发展,图书情报界对 MARC 格式不断作出改进的同时,在 MARC 中增加了相应的字段来揭示电子资源和超链接,其中最常用的就是 856 字段。将电子资源整合到 OPAC 系统就是通过定义 856 字段来实现的。

856 字段即电子资源地址与检索字段,它包括定位与检索电子资源所需要的信息。当文献资源或其子集可以通过参赛过电子方式获取,可在书目记录中启用该字段。在 CNAMRC 中,856 字段著录格式为:字段指示符 1 表示访问方式。0 电子邮件 (Email); 1 文件传输协议 (FTP); 2 远程登陆 (Telnet); 3 拨号放网 (Dial-up); 4 超文本传输协议 (HTTP)。指示符 2 表示未定义。

常用的子字段有:\$ u 统一资源标识 (URI),用于建立电子资源的统一资源地址 (URL);\$ z 公共附注,用于标识电子资源地址相关的附注。

通过定义 856 字段可以实现馆藏信息资源的简单整合,不过 856 字段仅仅只能描述连接电子资源地址,而不能对电子资源的元数据进行相

关描述,而且无法自动关联和快速删除和有效连接维护。

3 基于数据逻辑集成的电子资源整合

基于数据逻辑集成的整合模式是借助中间件技术或遵从异构数据库互操作协议(如: Z39.50, OA I-PMH, OpenURL)将多个异构的数字资源系统集成成为一个统一的数据视图^[5]。中间件技术是目前比较流行的实现异构数据统一检索方法之一。它是在异构数据库的基础上,借助中间件构造一个虚拟的、逻辑的全局数据模式,用户查询请求先提交至中间件,中间件将查询分解成对不同原始数据源的独立访问请求,通过标准(ODBC,DBC)或非标准的数据访问接口(API)对原始数据源进行实时检索访问,在得到各数据库返回结果后进行综合并映射转换为统一的格式显示给用户,使得用户可以把集成数据源作为一个统一的整体。

此种方式的特点是,中间件层并不存储具体的数据,只存储所有数据的逻辑集成模式。因此无须改变原数据库数据的存储和管理方式,是一种集中式管理、分布式存储的数据集成模式。这种整合方式的优点主要体现在:实时性好,任何原始数据的更改都可以在用户查询时及时得到反映;原文获取可以直接通过中间件获得从而省去原文数据源的原文发布服务。缺点主要体现在:原始数据源必须提供访问接口;由于各数据源的速度问题而影响用户得到结果的时间;中间件的开发较为复杂,获取各数据源后的数据整合策略和技术需要合理规划。

异构数据库资源整合是对不同数据库中相同学科专业的数字对象进行优化整合,也就是把在不同时间、用不同技术开发的、具有不同内容和形式的数字资源系统集成起来,屏蔽了异构资源系统间的差异,呈现给用户的是一个统一检索界面,用户可一次集成检索多个数字资源系统。整合后的数字资源具有相同的组织结构和组织功能,提供统一的用户界面和共同的检索方法。这种信息资源整合模式的主要实现方式是异构数据库统一检索平台(Union Search Platform,简称 USP)。USP是一个用于同时调用多个数据库和搜索引擎、进行资源检索的软件系统,它可以帮助用户同时在多个数据库中进行检索,同时得到多个数据库的结果。用户所看到的

是一个统一的界面,在对数据库进行选择后,进行单库或跨库检索,而不必分别进入不同的数据库接口,摸索掌握不同的数据库检索模式,使用户感觉就像在使用一个数据库进行检索操作一样。

异构数据库整合的常用技术是构建中间层,当用户提出检索请求后,其请求被交给服务器端的一个 Agent 程序。Agent 针对不同的数据库,将用户请求转化为符合其规定的格式,然后将请求发送到各数据库。在得到数据库的返回结果后,Agent 再将不同数据库的结果转化为统一的格式,并发送到浏览器端显示给用户。

异构数据库整合模式在图书馆中已经有所应用,如北京大学图书馆把目前拥有的 200 多个数据库、近 2 万种电子期刊和 10 余万种电子书整合在统一检索平台上,提供异构系统的跨库检索服务。用户可按学科、数据库名称、文种同时检索多个数据库的资源。上海数字图书馆项目也实现了对异构数据库信息资源的整合,该项目启动工程组织了丰富的馆藏文献,包括数以万计古籍、民国图书、地方文献、科技报告、中外期刊、音响资料、历史照片等信息资源,按照读者需求和文献特征形成九大系列,即上海图典、上海文典、点曲台、古籍善本、科技会议录、中国报刊、民国图书、西文期刊目次、科技百花园等。上海数字图书馆项目运用先进成熟的数字技术和网络技术,采取统一的界面、统一的软件、统一的管理,充分考虑当前需要、适应资源共享和可持续发展的目标,实现远程、快速、全面、有序、智能、特色六大服务优势。

现阶段数字资源主要包括图书馆 MARC 格式目录库所反映的现实馆藏资源、光盘数据库、网络数据库、电子虚拟馆藏以及一些图书馆自建的数据库和特藏资源数据库等,目前数字资源整合的重点在于将多种形式、多种类型、分散异构的信息资源有机的结合在一起,使用户能够在统一的数据存取模式下,通过统一的用户界面完成对不同数据库和网络资源的检索。

六、结束语

非书资源的整合和服务要以人为本。其价值因人因时因地而异。因为用户是这些非书资

源整合成果的利用者和评判者, 他们的需求千差万别, 因此非书资源整合的深度与整合结果要切合用户需求, 让用户以最小努力就可从图书馆的非书资源整合和推送服务中得到所需的信息和知识。在网络空间, 我们可以应用新的技术、软件来研究用户的信息查询, 使得非书资源的整合和提供会更符合用户的个性化要求, 发挥信息的最大价值。资源越来越多, 用户可能也不需要检索所有的资源, 只有在个性化上体现整合的目标, 才能满足用户不断发展的需求。

参考文献:

- [1] 路莹, 李炎. 基于 OPAC 系统的图书馆资源整合 [J]. 中华医学图书情报杂志, 2007, 16(4): 26- 28
- [2] 庞跃霞, 曹丽娟. 高校图书馆馆藏目录整合方法探讨 [J]. 图书馆杂志, 2006, (4): 40- 42
- [3] 谢玲. 论现阶段数字资源整合的实用技术及发展趋势 [J]. 现代情报, 2007, (8): 100- 103
- [4] 吴松华, 季亚娟. 随书光盘数据库的创建与应用 [J]. 现代情报, 2004, (8): 127- 128
- [5] 彭译华. 数字资源整合技术在数字图书馆建设中的应用 [J]. 高校图书馆工作, 2007, (5): 9- 12

Construction of None- book Materials Management and Library Digitization Management System

YE Lan

(Library, Suzhou University of Science and Technology, Suzhou 215011, China)

Abstract Resources integration is an important fundamental project of a digital library. The ideal integration resources of digital information, which links all kinds of information transparently and seamlessly, allows people to use it as only one operating system working. Thus, it becomes a significant question for discussion that how to integrate the existed OPAC to work with the various kinds of digital resources, electronic resources and network resources in order to realize effective using. And the integrated resources will exert powerful capability of the computers and networks, which make the information research efficiency improve greatly. On the basis of management system choosing hardware collocating and managing resources processing and operating, this paper introduces practices of the digital resources and the service platform of Suzhou University of Science and Technology. It also proposes that it exists some problems of digital resources with the OPAC, and puts forward some methods to deal with.

Key words none- book materials; integration; OPAC; library

[责任编辑 张大勇]