



基于 OPAC 的 图书馆资源整合研究

周 晨 ◎ 主编



國家圖書館出版社
National Library of China Publishing House

基于 OPAC 的图书馆资源 整合研究

周 晨 主编



國家圖書館出版社

National Library of China Publishing House

图书在版编目(CIP)数据

基于 OPAC 的图书馆资源整合研究/周晨主编.

—北京:国家图书馆出版社,2013.7

ISBN 978-7-5013-5122-0

I. ①基… II. ①周… III. ①图书馆利用-研究

IV. ①G252.65

中国版本图书馆 CIP 数据核字(2013)第 115255 号

责任编辑:高爽 王炳乾

书名 基于 OPAC 的图书馆资源整合研究

著者 周晨 主编

出版 国家图书馆出版社(100034 北京市西城区文津街 7 号)
(原北京图书馆出版社)

发行 010-66114536 66126153 66151313 66175620
66121706(传真) 66126156(门市部)

E-mail btsfxb@nlc.gov.cn(邮购)

Website www.nlcpress.com→投稿中心

经销 新华书店

印刷 北京科信印刷有限公司

开本 880×1230(毫米) 1/32

印张 5.75

版次 2013 年 7 月第 1 版 2013 年 7 月第 1 次印刷

字数 150 千字

书号 ISBN 978-7-5013-5122-0

定价 38.00 元

前 言

OPAC,是 Online Public Access Catalogue 的简称,国内一般通行的译法为联机公共检索目录。OPAC 是用户从图书馆获取信息的最基本、最通用的途径,所提供的服务与功能是全面衡量一个图书馆服务水平的重要指标。OPAC 是图书馆最早建设的数字资源之一,其本身也是图书馆最基本的数字资源。作为一种揭示与利用馆藏资源最基本、最直接的手段,近年来随着计算机技术、网络技术和情报检索技术的发展,它已逐步突破了时间与空间的限制,在使用及服务上也有新的突破,尤其是 Web2.0 技术为 OPAC 的发展带来了新的机遇。将图书馆数字资源与 OPAC 书目查询整合在一起,为读者开辟更加快捷地获取文献全文的渠道,具有重要的现实意义。

进入 21 世纪以来,基于 OPAC 的图书馆资源整合作为资源整合的基本方式得到了图书馆界广泛的重视,迅速发展。在理论研究方面,学术界对电子资源编目、元数据方案、资源整合技术、资源整合方法等进行积极的理论探讨。从实践层面上看,国内外图书馆关于基于 OPAC 的资源整合的实践主要集中在 4 个方面:第一,基于 856 字段的整合;第二,基于元数据仓储的整合;第三,基于开放接口的分布式整合;第四,基于 URL 传递检索式的整合。虽然近年来基于 OPAC 的图书馆资源整合研究和实践发展迅速,但从宏观上看,多数研究还缺乏深度和广度,尚未形成完整的理论体系。由于缺乏必要的理论与方法指导,各图书馆的整合实践带有不同程度的随意性和盲目性。

基于 OPAC 的图书馆资源整合,是资源整合实践发展的需要,来自实践又必将指导实践。今天,人们逐渐认识到,数字资源存在着不

同于传统资源的复杂特点,它直接影响对数字资源的获取与利用,导致数字资源组织与传统资源组织的诸多不同。同时,人们也清楚地意识到,数字资源整合实践需要理论的指导,因为在迅速发展的数字资源整合实践面前,传统的资源组织理论与方法有其不适应性,需要不断完善和发展。因此,有必要对国内外图书馆十多年来的资源整合实践进行回顾和总结,从宏观上对整合理论进行系统研究,提炼出既具普适性又有针对性的原理和方法,从而给图书馆资源整合实践提供指导与借鉴。

目 录

前 言	(1)
第一章 资源整合的内涵	(1)
第一节 资源整合的概念	(1)
第二节 资源整合的模式	(2)
第三节 资源整合的技术	(4)
第二章 国内外图书馆基于 OPAC 的资源整合实践	(9)
第一节 美国国会图书馆	(9)
第二节 澳大利亚国家图书馆	(20)
第三节 OCLC WorldCat	(31)
第四节 北卡罗来纳州立大学图书馆	(41)
第五节 香港大学图书馆	(49)
第六节 清华大学图书馆	(59)
第三章 基于 OPAC 的馆藏资源整合理论研究	(70)
第一节 OPAC 概述	(70)
第二节 基于 OPAC 的馆藏资源整合的必要性和可行性 ..	(76)
第三节 基于 OPAC 资源整合的内容	(79)
第四节 基于 OPAC 资源整合的层次	(88)
第五节 基于 OPAC 资源整合的方法	(99)

第六节 基于图书馆 OPAC 资源整合的技术标准	(107)
第七节 电子资源的著录规则	(119)
第四章 国家图书馆基于 OPAC 的资源整合实践研究	(129)
第一节 国家图书馆数字资源概述	(129)
第二节 国家图书馆书目数据现状	(130)
第三节 国家图书馆 OPAC 资源整合原则	(131)
第四节 基于 OPAC 的电子图书资源整合	(134)
第五节 基于 OPAC 的电子期刊资源整合	(144)
第六节 基于 OPAC 的自建数字资源整合	(151)
第七节 基于 OPAC 的网络资源整合	(157)
第五章 下一代 OPAC 与资源整合展望	(160)
第一节 OPAC2.0 阶段	(160)
第二节 OPAC3.0 时代	(164)
结 语	(168)
参考文献	(170)

第一章 资源整合的内涵

第一节 资源整合的概念

数字资源整合是指运用网络技术和应用软件把分散的信息资源连接起来构成集成化系统,以实现数字资源的集中管理和跨平台、跨数据库检索。换言之,就是对不同的数字资源实现无缝集成,并依据一定的需要,对各个相对独立的数字资源系统中的数据对象、功能结构及其互动关系进行融合、类聚和重组,使其重新结合为一个新的有机整体,形成一个效能更好、效率更高的新的数字资源体系,让使用者通过一个统一的检索界面查询到几乎所有的馆藏数字资源,以此来满足不同用户的个性化需求。数字资源整合是对日益增加的数字资源进行有序化管理,注重对数字资源的整体规划与管理,按资源的不同属性进行分类组织,重构信息资源的整体架构,整合图书馆全部数字资源,为用户提供更优质的服务。

从知识组织角度来说,数字资源整合是指相对独立的数字资源实现无缝链接进而产生新质的一种知识组织方法。它是一种数字资源优化组合的存在状态,其整合程度直接关系到数字资源能否被高效吸收与利用。而知识组织基于信息组织,是信息组织的深化。因而将知识组织作为数字资源整合的理论基础是合适的,也是科学的。知识组织是把知识客体中的知识因子和知识关联表示出来,以便人们识别和理解知识。对于数字资源整合而言,知识组织原理就是对处于不同地方的数字资源进行精简、提取,发现其中的有用知识单元,按知识管理规则将其集成、优化,而后提供给用户一个统一的数据平台。从而实现在

已有的信息基础上的信息增值。正确把握数字资源整合的内涵,有助于我们对信息整合做出科学合理的判断,并制订行之有效的策略。

第二节 资源整合的模式

近年来,在理论探讨和实践中产生了不同类型的数字资源整合方式。根据图书馆数字资源加工处理的程度及技术方法,可分为三个层面的整合,即数据整合、信息整合和知识整合^①。所谓数据整合是对异构资源系统中的异质、异类的数据在逻辑上或物理上的集中,提供统一的表示和查询,以解决多种异构数据资源的互联与共享;所谓信息整合,主要指通过某种机制或标准,对不同性质、不同来源和不同格式的数字资源进行描述与连接,使相对独立的资源实体产生联系,实现数字资源的全方位整合和统一获取;所谓知识整合,是以信息整合为基础,以知识组织体系为支撑,组织资源知识结构中概念及概念关系的一种整合方式。这 3 种方式的数字资源整合各有侧重,实际的数字资源整合系统是以某种方式为主并兼有其他整合方式的混合体。

本书将按照涉及的对象及整合的过程,从实际操作角度,把图书馆数字资源整合分为 3 种模式:基于信息链接的整合、基于跨库检索的整合、基于元数据仓库的整合。

一、基于信息链接的整合

基于链接系统的数字资源整合主要是借助超文本的链接特性,链接起文献的相关知识面,使有关的数字资源链接在一起,联成一个有机的整体。目前,主要有 3 种不同的链接方式:封闭式静态链接系统、

① 马文峰,杜小勇. 数字资源整合:理论、方法与应用[M]. 北京:北京图书馆出版社,2007.

开放式静态链接系统、开放式动态链接系统。封闭式静态链接系统是指所有的链接系统只存在本地数据之间,并且所有的链接都是事先计算,并通过存放于专门数据库的链接记录来表达。开放式静态链接系统采用的技术方法是资源拥有者提供链接对象的 URL 或标识符,比如全文出版商向文摘出版商提供期刊文章的标识符和 URL,文摘出版商将其嵌入自己的文摘索引数据库中,实现两者之间的链接。开放式动态链接系统不同于封闭式静态链接系统与开放式静态链接系统,这两个系统的链接源与链接对象都是事先固化好的,只能处理一对一的链接,而对于出现的新链接对象、一对多情况下的选择性链接等则不能发挥作用,开放式动态链接主要是指基于 OpenURL 开放式统一协议建立的链接系统,实现基于 Web 的学术信息环境上下文相关链接传递服务。

二、基于跨库检索的整合

跨库检索是在同一检索平台下实现对多数据库的同时检索,提高信息资源检索的效率。跨库检索系统的数字资源整合可以分为两层:第一是整合检索界面;第二是实现数字资源系统间的异构检索系统。清华同方异构数据库实行的是 USP 网络资源共享平台,可以同时调动多个数据库和搜索引擎进行检索的软件系统,同时能获取多个数据库的结果,这样能避免用户逐个登录数据库,重复输入检索条件的麻烦;艾利贝斯公司的 MetaLib 系统也是商业跨库检索系统的典型代表,能将纸质文献、电子文献通过联邦检索的形式整合在一起,实现同时获取;还有 CALIS 数字图书馆资源统一检索系统和国家科学数字图书馆跨库集成检索系统都是这种整合方式的代表。目前,统一资源检索平台在技术上还存在一些不足,但是它的出现已经在很大程度上满足了用户的需求,这也是目前数字图书馆能为用户提供的实现资源整合的典范方式^①。

① 王春玲. 数字图书馆资源整合探讨[J]. 中国报业, 2011(20): 81-82.

三、基于元数据仓库的整合

数据仓库不同于一般的数据库,它是按照一定的要求对多个异构数据源所进行的有效集成,它的主要特征是数据组织的主题性、数据存储的集成性、数据格式的一致性、数据的相对稳定性和数据内容的更新性。基于元数据仓库的整合方式通过对来自异地、异构的数据源或数据库的元数据进行抽取、解析、净化、转换、过滤、整合,使分散的、不一致的数据转换成公共数据模型并被集成到数据仓库中,用户通过对本地数据库的访问实现多个异构数据库的一次性检索。基于元数据仓库整合的主要特点是:元数据集中,对象数据分散,解决了分散的异构数据源间的不同和差异,为用户提供一个单一、透明的数据使用环境,查询速度快,数据质量高,但数据集成消耗大量的存储空间,获取数据有一定的时滞性。国家图书馆的文津搜索系统采用的就是这种整合模式,将自建的、外购的以及其他渠道获取的数据集成到一个大的数据仓库中,实现用户的一站式检索;艾利贝斯公司的 Primo 系统是商业元数据仓库整合系统的典型代表。

第三节 资源整合的技术

一、基于信息链接的整合技术

基于信息链接的整合技术主要包括静态链接技术和动态链接技术。

1. 静态链接技术

静态链接的基本特征是将链接属性直接写为某一固定 URL,用户直接点击就可链接到对象。可分为封闭式静态链接和开放式静态链接。这两种链接状态是都是固定的,容易产生死链。封闭式静

态链接的逻辑控制主要由链接者本地实施,链接源也存在本地。比较常见的是数据库商内部的资源链接。例如许多信息服务商将自己的文摘索引数据库、全文期刊、文后参考文献、全文文章链接起来。开放式静态链接技术主要是逻辑控制在本地,链接源来自于外地。应用最常见的是图书馆内部 OPAC 书目数据库与其他类型数字资源的整合,在 MARC 记录里增加 856 字段,用于记录被著录的数字资源的存取地址和存取方式。不同的出版商之间的数据如果严格按照某种协议描述数据和传递数据,也能实现这种开放式的静态链接。

2. 动态链接技术

所谓动态链接是指根据用户的需要,根据一定规则计算链接路径和进行链接,既可对用户链接前刚出现的链接对象或位置予以链接,也可在链接计算规则中嵌入选择规则实现选择性链接,例如基于 OpenURL 技术的 SFX 就是动态链接的典型范例。但是,这种系统的规则描述难度较大。此技术克服了静态库先天上的不足,促成共享库(动态链接文件)的出现。共享库是一个目标模块,运行时,可以加载到任意存储器地址,并在存储器中和一个程序链接起来,这个过程叫动态链接,是由动态链接器完成的。

二、基于跨库检索的整合技术

目前国内外基于跨库检索的整合技术主要有以下几种:

1. 公共网关接口技术(CGI)

公共网关接口(Common Gate Interface, CGI)是 Web 服务器与 CGI 应用程序之间进行通信的协议标准。CGI 应用程序可以完成客户端和服务器的交互操作,它允许用户根据需要采用各种语言去实现无法用 HTTP、HTML 实现的功能。CGI 的工作原理是基于 HTTP 协议。HTTP 协议是建立在 TCP/IP 协议之上的“无连接”协议,每次连接只

处理一个请求。在服务器上,运行着一个守护程序对端口进行监听,等待来自客户的请求。当一个请求到来时,将创建一个子进程为用户的连接服务。一个 CGI 程序可以从数据库服务器获取数据,将其转化为 HTML 文档后发送给浏览器。另外,它也可以将从浏览器获取的数据发送到数据库中。

2. 开放式数据库互连技术(ODBC)

开放式数据库互连(Open DataBase Connectivity, ODBC)是由微软公司提出的一个用于访问数据库的统一界面标准。它通过使用相应平台上和所需数据库对应的 ODBC 驱动程序与应用程序的交互,实现数据库操作,而避免程序直接调用数据库,实现了数据库的独立性。绝大部分数据库厂商都为产品提供了 ODBC 接口或 ODBC 支持,包括 Access、SQL Server、Oracal 等。ODBC 的基本思想是为用户提供简单、标准、透明的数据库连接的公共编程接口。

3. 中间件(Middleware)技术

中间件位于异构数据库系统(数据层)和应用程序(应用层)之间,它的主要作用是处于上层的应用程序提供运行与开发的环境,帮助用户灵活、高效地开发和集成复杂的应用软件,为访问集成数据提供统一数据模式和数据访问的通用接口,用来实现异构数据库的互操作,其实质是为异构数据库提供一个高层次检索服务。中间件还是一种支持分布式应用的重要组建框架结构,分布式应用借助中间件在不同的系统之间共享资源。中间件技术包括基于互操作开放协议的跨库访问技术及分布式搜索技术,如 OAI 协议、Z39.50 协议、Web Service 等,关于这些内容以下章节将作详细介绍。

三、基于元数据仓库的整合技术

基于元数据仓库的整合技术主要包括:元数据标准框架(Metadata Framework)、元数据映射(Crosswalk)和元数据注册(Registry)。

1. 元数据标准框架(Metadata Framework)

元数据标准框架是一个参考模型(Reference Model),是抽象化的元数据,在各个元数据标准之上提供一个高层次、概念性的结构,规定了元数据的功能、数据结构、格式设计、方法语义、语法规则等多方面的内容。元数据编码传输标准(METS)就是一种元数据标准框架,它是专门针对数字图书馆对象的管理而开发的。METS将元数据划分为描述性元数据、管理性元数据、结构性元数据,其最重要的贡献就是对结构性元数据的强调。一个数字对象可能由多个文件构成,METS标准中的结构性元数据说明了数字对象中的各个文件是如何构成统一有机的数字实体的,从而将数字对象的各个组成部分封装成一个整体。这种结构性信息支持数字图书馆对数字对象的管理,同时也使得各数字图书馆之间的数字对象互换变得更为方便。METS标准采用XML DTD作为置标语言,可以同时采取多种元数据标准进行标引,只需要说明所使用的是哪种标准即可。

2. 元数据映射(Crosswalk)

除了元数据标准框架之外,元数据映射(Crosswalk)也经常被用来映射不同元数据标准之间的元素、语义及标记等。元数据映射使得一种元数据可以被另一种元数据标准使用。这种映射关系的兼容程度取决于两种元数据标准的相似程度、元数据标准元素的多少(元素的粒度)及内容规则。元数据映射对OAI的实施很关键,一个文档需要将本地的元数据格式映射成简单DC(DC的15个核心元素如题名、主题等),才能被发现和收割到中心存储库中。目前有很多工具可以完成通用元数据标准之间的映射,如MARC与DC、MODS与MARC等。

3. 元数据注册(Registry)

元数据注册是对元数据的定义信息、置标、映射、应用等规范进行发布、登记、管理和检索的一套机制。目的是支持开放环境中元

数据规范的发现、识别、调用以及在此基础上的元数据映射、挖掘和复用^①。如:DCMI 为 DC 元素开发了注册中心;美国环境保护局的环境数据注册中心,提供 EPA 数据库中使用的数以千计的数据元素的信息。

① 夏翠娟. 数字图书馆的语义互操作及其实现[D]. 上海:华东师范大学,2005.

第二章 国内外图书馆基于 OPAC 的资源整合实践

随着信息技术的飞速发展,特别是 Web2.0 和 Library2.0 的应用,图书馆从以资源为中心的时代进入以用户为中心的新时代。图书馆在迎来新机遇的同时,也面临了新的挑战,用户需求不断提高。近年来,各类型图书馆的资源建设都呈现迅猛发展的态势,资源类型结构发生了巨大变化,资源类型正在向多元化发展。电子图书、电子期刊、各类网络资源、文摘索引数据库、全文数据库以及自建的各种馆藏资源库等层出不穷,使原来的实体资源占据图书馆绝对主导地位发生了动摇,新时期的图书馆已经成为传统图书馆与数字图书馆结合的复合型图书馆。

图书馆由于受到技术及综合实力等因素的影响,从国家图书馆、公共图书馆到大学图书馆,其 OPAC 整合的状况参差不齐。本章以具体案例对各类型图书馆的 OPAC 整合实践进行了调研与分析。

第一节 美国国会图书馆

美国国会图书馆(Library of Congress, LC)建于 1800 年,至今已有 210 多年的历史。国会图书馆不仅是国会的研究支撑,还是美国的国家图书馆。国会图书馆对世界各类型文献均有收藏,数量多,文种、门类齐全。

一、门户网站

美国国会图书馆的门户网站提供对本网站资源的集成搜索,资源类型包括图书馆目录、数字资源以及网络资源,涵盖馆藏实体资源、美

国记忆、特藏、全球网关、托马斯法律数字图书馆和老兵历史等各种资源的检索。

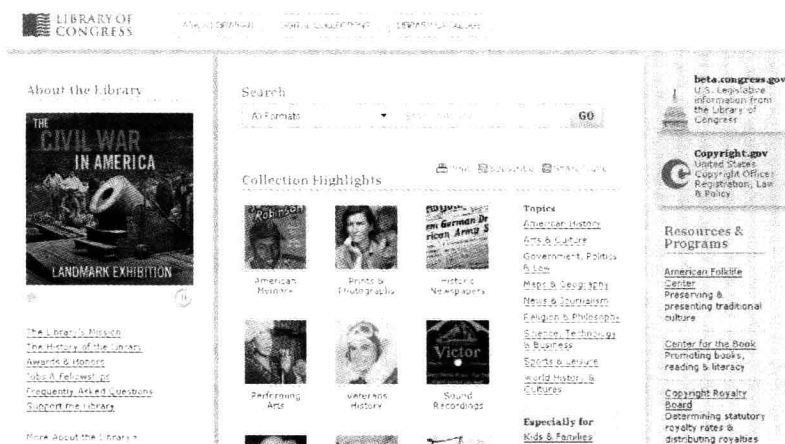


图 2-1 美国国会图书馆门户

门户页面的中间偏上方就是简洁的 OPAC 资源检索框, 用户可以根据需要检索国会图书馆所有格式的馆藏资源。也可以从左边的下拉框中选择音频、图书、电影/视频、立法、手稿、地图、乐谱、报纸、期刊、照片/打印稿/绘画、软件/电子资源、网页、3D 对象等不同类型的, 进行更精确的检索。如图 2-2 所示。



图 2-2 美国国会图书馆门户检索界面