

图书馆数字资源整合浅谈

邵燕, 宋文

(中科院文献情报中心, 北京 100080)

【摘要】论述数字资源整合的必要性, 结合实例重点介绍现有的几种整合方式, 并就其内在联系和发展方向进行探讨。

【关键词】数字资源整合; 跨库检索; OPAC; SFX; 资源导航

【中图分类号】G253 【文献标识码】A 【文章编号】1002-1167(2005)06-0187-03

“我们淹没在数据的海洋中, 却又忍受着知识的饥渴”, 用这句话来形容目前网络环境下人们面临信息爆炸的窘境真是非常的贴切。面对如今数字信息资源的迅猛发展以及用户日益加大的信息需求, 图书馆作为一个专门搜集、整理、保存并传播信息的机构, 也在不断的寻求如何更有效地组织和管理数字资源的途径, 寻求如何将这些不同内容不同形式的数字资源整合在一起, 实现资源、服务的一体化, 从而为用户提供最大的便利。

下面我们就针对图书馆数字化信息资源整合的相关问题做些探讨。

1 图书馆现有数字资源的服务存在的问题

数字资源也即数字化信息资源, 是包括所有以电子数据形式, 把文字、图像、声音、动画等多种形式的信息存放在光、磁等非印刷介质的载体上, 并通过网络通讯、计算机终端等方式再现出来的信息资源^[1]。目前大多数图书馆拥有的数字化信息资源, 不论是从内容还是形式上来说都非常的丰富。不仅包括传统的一次文献, 如图书、期刊、报纸、标准、专利产品等全文数据库; 还包括各种二次文献, 如文摘、题录和书目数据的数据库和相关系统。较之传统资源, 数字资源具有通用性、开放性和便于共享性。这是数字信息资源的巨大优势。然而在其为图书馆和用户提供了很多便利的同时, 也带来了很多问题。

1.1 数字信息资源与图书馆书目系统之间的分离

许多图书馆购买了很多的数字资源, 如电子图书、电子期刊等, 提供给用户使用, 但是在图书馆的书目系统中却检索不到相应的书目信息。还有些是购买了大量的国内外网络数据库资源, 如方正数字图书馆中文电子图书库、Elsevier、ACM等数据库, 但是没有包含在图书馆的馆藏目录或期刊目录里。不同的数据库覆盖的范围有时相差很大, 因此读者使用时为保证查全往往要在不同的库中反复检索。

1.2 不同类型数据库之间缺乏关联。

比如有些数据库仅提供文摘, 有的数据库能够提供全文或者引文。用户在使用时可能会出现这样的情况: 在文摘数据库中检索到篇名信息后想获取原文, 于是利用篇名信息进入 OPAC 进行重新检索, 得到相关馆藏信息, 如果不在本馆就还需要发出原文传递请求; 或者是, 用户查找到了某相关主题文献后, 想得到其引文, 就需要重新检索或者到别的库获取引文; 有时候在不同的全文或者文摘数据库中获取的结果之间存在着交叉重复现象, 用户要想获取全面的信息, 就要自己进行去重筛选工作。类似的情况对用户来说是很不方便的, 可能还会耗费大量的精力。

1.3 不同来源数据库之间存在很大差异

数字信息资源有许多的来源, 出版社、代理商、数据库公司等, 不同的机构开发的数字资源在内容、技术和形式上都有各自的特点, 不同的编码结构和表达方式造就了不同的描述和组织标准, 进而导致了检索途径和方法的差异。此外, 不同的数据库使用的检索软件也不同, 使得数据库的检索界面风格各异。甚至彼此之间由于竞争, 在数据库设计上还具有排他性, 这不仅给用户带来不便, 对图书馆的管理也造成了很大的困难。

2 数字化信息资源整合的必要性和原则

虽然现今拥有的资源存在这么多的问题, 但因为已经倾注了这么多的人力物力, 所以抛弃现有系统重新建设一个新的综合系统是不经济也不现实的。那么, 现在现有条件下, 该如何解决上述诸多问题呢?

对已有的数字资源进行有效整合, 建立各数字资源间的有机联系, 以统一友好的界面向用户提供服务, 才是唯一也是切实可行的途径。

数字资源整合就是将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。

数字资源整合除了要遵循一些一般性原则(如实用性原则、用户至上原则、系统性原则等)以外, 还要具备一些适应数字化网络环境的特定原则^[2]:

整体性原则。要保持各信息资源对象、功能的完整性;

标准化原则。要实现包括数据格式、描述语言、通讯协议, 数据库管理软件等的标准化;

优化性原则。要使资源得到合理组合, 取得最好的组织结构和组织功能。

3 数字化信息资源的整合方式

许多人都将数字资源的整合方式简单归为几类, 分别讨论。笔者认为, 目前存在的几种不同的数字资源整合方式其实并不存在很明显的分界, 追根溯源, 它们其实是存在着内在联系, 其产生或者说存在是有内在逻辑关系的。因此, 本文不是要从类别上对其进行区分, 而是试图从其解决的不同问题以及发展的相继相容方面来做些探讨。

3.1 跨库检索 (Federated Searching)

图书馆购置的数据库, 最开始是一个个独立的数据库, 不同学科, 也不同类别, 这对于用户使用来说非常不方便, 因为要保证查全率就得在不同的数据库之间穿梭, 浪费了大量的精力。这个时候就出现了跨库检索, 这是最初期也是最简单的数字资源整合方式。

跨库检索的工作原理是：当用户提出检索请求后，其请求被交给服务器端的一个称为“智能搜索器”的程序。“智能搜索器”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，“智能搜索器”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。

世界上最早的也是最大的联机检索系统 DIALOG 对其所购买的 450 多个异构数据库进行的整合就非常成功。这些数据库由不同的开发商提供，有书目、摘要、索引、全文等各种不同形式，并且检索方式也各不相同，但是 DIALOG 对这些数字资源进行了很好的整合，提供了统一的检索方法，使用者可以同时打开多个甚至全部数据库进行同样的检索，并且可以进行结果的去重，这一技术上的突破大大便利了用户的使用。

国内做得比较好的跨库检索系统比如清华同方的异构数据库统一检索平台 (USP, Union Search Platform)，就是一个用于同时调用多个数据库和搜索引擎进行资料检索的软件系统。USP 目前支持 30 多种国内外常用的搜索引擎或数据库，并且在不断增加。

跨库检索方式同时在多个数据库中检索并同时得到多个数据库的结果，避免了用户逐个登录数据库、输入检索条件的麻烦，使用方便、快捷。可以这么说，跨库检索方式从技术上屏蔽了图书馆的不同数据库之间检索方法和界面的差异。

但是，一个图书馆的馆藏毕竟有限，那么当读者在馆内查不到需要的资料时，还需要到别的图书馆或者信息机构进行查询，这对于用户来说是不方便的。另外，图书馆的馆藏不止是电子资源，还有大量的印本馆藏、多媒体资源等，如何将其结合起来提供给读者使用？

因此，跨库检索只做到基于数据库的检索和界面的整合是远远不够的。

3.2 利用图书馆 OPAC 系统进行资源整合

OPAC (联机公共检索目录) 是一种通过因特网对馆藏资源进行检索的工具，是网络环境下用户利用图书馆的入口，也是数字化图书馆最基础的一项建设。OPAC 就相当于传统的卡片式目录系统，因此，也应当是反映图书馆全部馆藏的一个检索窗口，使用户可以通过统一的界面，一次性检索即可获得有关主题的不同载体的馆藏信息。同时，由于出版物数量的剧增和图书馆有限经费之间的矛盾，使得任何一个馆都不可能聚齐所有的文献，只能根据本馆服务需要来选择收藏范围，因此，在本馆馆藏不能满足用户需要时，如果能指引用户利用其他馆馆藏，就可以通过馆际互借等形式，实现资源共享^[3]。

因此，基于 OPAC 进行资源整合，就可以很好的解决上述跨库检索无法实现的功能。而这两个方面可以归结为馆内资源整合和馆外资源整合。

馆内整合。在传统 OPAC 的基础上，提供多种形式的资源。利用《国际标准书目著录》、《英美编目条例》中对连续性出版物、声像资料、电子出版物等的著录细则，在 MARC 记录中相应字段进行揭示或链接 (比如 USMARC 中的 007、530、776、856 字段，CNMARC 中的 230、337、856 等字段)，实现 OPAC 书目信息与多种形式资源之间的整合。如中国人民大学图书馆 OPAC 系统就将图书、期刊 (包括印本和电子形式的)、光盘、音频、视频、图像、网络资源等，按照著录规则和 MARC 格式的要求，集合在一个数据库中。

馆外整合。也就是本馆与其他馆 OPAC 系统之间进行的整合，当前较多采用 Z39.50 协议来完成。通过执行 Z39.50 协议，首先将需要整合的图书馆书目系统映像成自己的专用模型，再根据本馆的

要求建立统一的检索接口，从而将不同平台上的异构 OPAC 数据库聚合在一起。如上海图书馆所采用的“Wcbpac 检索系统”，可在统一检索界面上通过著者、题目、主题、关键词、标准书号等途径，检索上海交通大学图书馆、华东师范大学图书馆、复旦大学图书馆、上海图书馆的 OPAC 书目信息，而不需要在各个图书馆的 OPAC 界面间来回切换。这种整合方式只是一种叠加，不打乱各自的结构，不影响各自的检索结果。

基于 OPAC 系统进行的资源整合，对不同形式的信息资源进行揭示或者提供链接，并且实现了不同馆之间的资源共享。但是，其对资源的链接都是静态的，一旦目标资源发生了变化，链接点就要进行更新维护，这是一项非常大的工程，因而不可避免会产生很多无效链接。另外，馆外整合实现的是不同馆检索界面的整合，但是馆内购买的不同来源数据库之间在数据格式及通信协议方面也都存在这差异，而简单的 OPAC 界面整合对此是无能为力的。

这就促成了基于软件进行深层的信息资源整合的出现。

3.3 利用 SFX 进行资源整合

SFX 即 Special Effects 的缩写，是 Ex Libris 公司的基于开放的统一资源定位器 (OpenURL) 标准的上下文相关 (Context-Sensitive) 参考链接系统。SFX 可以把不同来源和不同通信协议的信息完全融合，使不同类型、不同格式的数字资源实现无缝连接。它为图书馆信息资源整合提供了灵活性和可选择性。图书馆可以从大范围的信息提供商那里选择合适的内容，根据需要来实现这些内容的互连，并且为最终用户连接到合适的服务。同时，在动态的连接环境中，SFX 是在用户提交请求的过程中创建的，而其强大的计算法则可确定连接资源的可获得性。对于参与 SFX 框架的每个服务提供商，在 SFX 内都有一个单独的全局模板，如果其站点有所变化，该全局模板随之改变，这样，即使地址发生改变，分布式的全局连接模板也能识别，并在使用 SFX 的团体内进行信息更新^[4]。

具体到图书馆的资源，利用 SFX 实现的整合可概括为^[5]：

OPAC 与数据库之间的动态链接。用户通过 OPAC 进行检索，结果显示有关此主题的所有信息：印本馆藏显示地点，电子期刊提供动态生成的链接地址，电子书点击则进入相应的电子书系统浏览全文。

引文数据库、二次文献数据库与全文数据库之间的链接。二次文献数据库检索功能强，用户可以根据检索到的文摘、索引信息直接获得想要了解的论文全文。而检索结果的引文资源对研究者的借鉴价值之大，已是毋庸置疑的，所以进行引文与其相应的全文资源进行链接，将给用户提供很大的便利。

数据库和 WEB 资源的链接。这其实是向外部进行的一种扩展链接。用户在检索到某主题的结果后，可通过相关链接指引到相关主题的 WEB 资源，如通过搜索引擎搜索的结果或者相关的网站信息。这样用户就可以一站式获取多种信息。

图书馆长期以来都是分类和主题两种途径并列管理资源提供服务，因此对于资源的整合也少不了分类组织方式的应用。前面的几种整合很大程度上都是从信息的主题相关性方面切入的，该如何将组织结构很严密的分类体系应用到图书馆庞大纷繁的数字资源中呢？这就需要利用资源导航进行数字资源整合。

3.4 利用资源导航进行数字资源整合

资源导航是借鉴学科导航的概念，但是不但需要从学科角度出发对数字资源进行管理，同时还应该从数据库或者文献类型方面进行分类、描述，提供链接和检索等相关服务。将各数据库包括的文献分离重组，按照学科、主题、字顺等再次序化，组成多种导航方

式;或者将数据库按照收录的文献类型分为全文、文摘、学位论文、会议录等不同类型,提供用户浏览使用。

比如国家科学数字图书馆 CSDL 就在其主页上实现了资源导航功能。在其页面上,我们可以看到首先是网络数据库按照收录类型分为全文、文摘和其他类,每类下面是数据库列表,点击就进入到相关的数据库页面。其次是对网络电子期刊按学科进行分类(数学、物理、化学、地学、生物、农林、医学、工程、环境、管理、社科、图情),同时又根据情况在某些大类下设置了二级类目。同时还提供电子期刊首字母浏览和按出版商/数据库浏览。

此外,CSDL 还整合了因特网上各学科领域的文献信息资源和服务,对各个学科领域网络资源提供权威可靠的导航,目前包括化学、资源环境、生命科学、数学、物理和图书情报6个学科门户。

资源导航将分类组织方法引入电子资源的管理,成功地由于数据库本身各种差异带来的不便屏蔽掉了,同时,经过结构化的资源列表更便于用户快速定位到自己所需的相关主题处,尤其适用于那些对所查找的内容没有明确概念的用户。当然,为了获得这样的效果,类目的设置(类别和标引的深度)就变得非常重要。

总的来说,跨库检索是从技术上为用户屏蔽各个数据库使用时的差异,而学科导航则是将用户检索数据库的精力转移到对内容的浏览上,这样便于更快找到所需数据库。一定意义上说,OPAC 是整合数字资源很重要的一环,通过设置馆藏书目记录相关字段进行资源整合,有利于图书馆从整体上对馆藏进行管理,并且将 Z39.50 协议和 SFX 等软件与 OPAC 结合在一起,相互取长补短,实现更强大的整合功能。比如,如果各个 OPAC 系统都支持 OpenURL^[7],则不需要建立庞大的联合目录,只需要在链接服务器上作相应的设置,就可实现对不同 OPAC 系统的统一检索;而 SFX 软件要想实现其对资源的动态链接等功能,数据来源就是图书馆的书目数据库。

上述方式内在的联系和互补性决定了我们在进行整合时,不应该只考虑其中某一种,而应该结合使用。同时,仅仅就资源进行整合还是不够的,因为图书馆的资源和服务是密切相关的,因此,只有在整合资源的同时,将相关服务与其绑定,才能真正解决用户的需求,为用户提供一站式信息服务。

4 结语

因此,要将上述整合方式结合起来,同时在馆际互借、参考咨询、个性化定制等服务与相应的资源之间建立灵活的联系,充分利用网络和馆内丰富的资源,解决用户的实际问题,从而也可以将图书馆员从烦琐的机械性劳动中解脱出来,从事更有意义的工作。就图书馆管理来说,还可以利用通过馆藏书目数据和用户使用情况等

实现对期刊的统计评价功能,指导馆藏建设和资源布局。目前比较有代表性的系统有:国外的比如 Jake 系统以及同一公司推出的新产品 1Cate^[8],还有 TDNe^[9]、t. Serials Solutions、JournalWebCite 等系统^[10],国内的有正在建设当中的有 CSDL 资助的中国科学文献数据库服务系统(Science China^[11])项目等。对于国内图书馆进行数字资源建设都是很有学习价值的。

图书馆在信息社会扮演着重要的角色,尽管现在出现了很多替代性工具和信息源,比如 google 和网志等,它们甚至对图书馆存在的意义有一定的颠覆性,但是,笔者认为,无论是那种信息源,重要的是能够真正的从用户需要的角度去考虑,切实满足用户的需求,这是最基本的指导原则。而经过整合的图书馆资源,无论是从质量还是从组织的深度上都拥有着其他资源无法比拟的优势,并且,二者未必就是矛盾的,图书馆管理系统还可以很好的利用那些工具和信息源,将其整合到自己的系统当中,从而提供更全面的服务。

目前国内的数字资源整合在理论和实践上都有了很大进展,但还需要我们以用户为本,积极的吸收和借鉴国内外先进的技术和系统,共同参与到数字图书馆建设当中去;真正做到为用户提供无缝链接的信息和一站式服务,创造一个以用户为中心的良好学习和科研环境。

〔参考文献〕

- [1] 王 辉,康美娟. 数字资源的整合探讨 [J]. 情报业务研究, 2004, (8).
- [2] 马大川,杨红平. 信息资源的集成整合研究 [J]. 中国图书馆学报, 2004, (3).
- [3] 胡 宁. 网络环境下 OPAC 系统资源整合 [J]. 上海高校图书馆情报工作研究, 2003, (1).
- [4] <http://www.exlibris.com.cn/2004.11-07>.
- [5] 李富玲,卢振波. SFX——信息资源整合新工具 [J]. 图书情报技术, 2002, (6).
- [6] <http://www.csdl.cn/2004-11-05>.
- [7] 林绮屏. OpenURL 与数字资源整合 [J]. 科技情报开发与经济, 2004, (7).
- [8] <http://www.openly.com/1cate/2004-10-19>.
- [9] <http://www.tdnet.com/2004-10-25>.
- [10] Michelle Sitko, Narda Tafuri, Gregory Szczyrbak, and Taemin Park. E-journal Management Systems: Trends, Trials, and Trade-offs. Serials Review 2002, (28): 176-194.
- [11] <http://sciencechina.net.cn/2004-11-15>.

A Brief Analysis of the Integration of Library Digital Resources

SHAO Yan, SONG Wen

(Library of Chinese Academy of Sciences, Beijing 100080, China)

Abstract: In the digital environment, information resource are produced at an amazing rate, which provides not only great convenience but also many problems. What should libraries do to their resources and thus provides users integrated resources and convenient services. This article tries to discuss some issues about the integration of digital resources.

Key words: digital resource integration; federated searching; OPAC; SFX; resource navigation

〔作者简介〕 邵 燕 (1981-) 女, 中科院文献情报中心研究生;

宋 文 (1961-) 女, 研究馆员, 中科院文献情报中心部门副主任。

〔收稿日期〕 2004-12-09