

·信息工作·

我国图书馆数字资源整合研究

黄传慧 (武汉大学信息管理学院 湖北武汉 430072)

李娟 (清华大学图书馆 北京 100084)

摘要: 文章分析了我国图书馆数字资源的现状,依据数据、信息、知识三者的关系以及对数字资源加工程度的深化,提出图书馆数字资源的整合目标层次是数据整合——信息整合——知识整合,并以整合目标层次为基础,分析了基于OPAC、基于跨库检索、基于资源导航、基于超级链接、基于知识本体的数字资源整合五种整合模式,在比较了五种整合模式的优缺点以后,指出基于知识本体的数字资源整合是未来整合模式的发展趋势。

关键词: 图书馆 数字资源 整合研究

中图分类号: G250.7

文献标识码: A

文章编号: 1003-6938(2009)04-0066-04

Research on Digital Resources Integration of China Library

Huang Chuanhui (School of Information Management, Wuhan university, Wuhan, Hubei, 430072)

Li Juan (Tsinghua University Library, Beijing, 100084)

Abstract: This article analyzes the status of digital library resources. Based on the relationship between data, information and knowledge and according as the deepening of processing for digital resources, this article puts forward that the target level of library resources integration is data integration - information integration - Knowledge Integration. Then, this article analyzes five integration models, which respectively based on OPAC, cross-database searching, resources navigation, hyperlinks and knowledge ontology. After comparing the advantages and disadvantages of the five integration models, this article points out that Ontology-based integration model is the development trend in the future.

Key words: library; digital resources; integration research

CLC number: G250.7

Document code: A

Article ID: 1003-6938(2009)04-0066-04

1 前言

二十一世纪的图书馆面临着信息环境的巨大变化,随着计算机技术、网络技术的飞速发展,图书馆馆藏结构发生改变,馆藏范围超出了印刷资料、缩微资料、视听资料等传统范围,延伸到各种数字资源。数字技术在图书馆的应用,使得图书馆数字资源日益丰富,种类越来越多,来源越来越广泛,但由于不同的数字资源分布在不同的数据库中,数据的组织、管理形式以及存储格式各不相同,不同的数据库构建方式和支持平台也不相同,产生了数据库间的异构性问题。各个数据库内容交叉重复,知识关联程度很低。当用户查找资料时,各个数据库检索界面和方法又不一样,用户必须掌握不同的检索软件使用方法,经常需要在不同的系统之间来回穿梭、重复操作,用户无法进行“一站式”的便

利检索,而且查准率和查全率都难以保证。要想改变这种状况,对图书馆数字资源进行整合是行之有效的解决办法,本文分析了数字资源整合的目标层次,探讨了数字资源整合模式的发展,在比较了几种整合模式的优缺点之后,指出基于知识本体的数字资源整合模式是未来发展趋势。

2 图书馆数字资源整合的目标层次

数字资源整合就是依据一定的需要,通过中间技术,把不同来源、不同通信协议的相对独立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,实现无缝链接,重新结合为一个新的有机整体。^[1]用户通过简单的方法同时检索不同的数据库,从开始检索到最终获得信息,都在统一的界面中进行,如同在一种信息资源系统中操作一样,满足用户从多个

基金项目:本文系国家自然科学基金项目“基于语义网络的数字图书馆应用模型研究”(项目编号:70803009)研究成果之一。

收稿日期:2009-02-15 责任编辑:魏志鹏

分布式数据源抽取和合并数据的查询请求,从而形成一个性能更优、服务更好的数字资源体系。在指导图书馆如何进行资源整合时,探讨数字资源整合模式很有必要,而分析数字资源整合的目标层次将使整合模式更有针对性。

分析图书馆数字资源整合的目标层次,不妨先看看读者对图书馆资源的需求层次。读者到图书馆最传统的需求就是文献资源的直接获取,实际上就是对数据的获取,信息技术的推动使得读者对图书馆的需求进一步提高,读者希望由直接的数据获取到信息的获取,随着知识经济的到来,读者不希望将更多的时间花在海量信息的查询和检索上,期望能以最短的时间和最方便的方式获取经过整理系统的知识,所以,数据——信息——知识,是读者对图书馆需求层次由低级到高级的逐步提升。先看看三者之间的关系,数据是信息的载体,信息是加工后的数据,能够减少不确定性,对使用者的决策有意义。例如,“明天会下雨”,这是一条数据,而当你说,“明天会下雨,我们取消户外活动”时,这条数据变成了一条信息。从对信息和知识的转化关系来看,知识是通过信息技术加工过的信息,信息通过实践活动产生经验转化为知识。知识是人们在改造世界的实践中所获得的认识和经验的总和,而信息是生产知识的原料或半成品,信息传入人的大脑,产生刺激,经过选择、加工和整理,组成新的信息系统,即新知识,知识是人类特有的信息,是信息的一部分。^[2]数据、信息、知识是图书情报研究的基本概念,也是图书馆整合重要的研究对象,因此我们认为,图书馆数字资源整合体系包括数据整合——信息整合——知识整合三个层次。

2.1 数据整合

上世纪90年代,图书馆数字资源在数量和规模上迅猛增长,形式异常丰富,数据库不仅包括光盘数据库和自建数据库,还包括大量联机数据库,如国内的中国学术期刊全文数据库、万方数据库、书生之家等,国外的EBSCO全文数据库、Web of Science引文数据库、BIO-SIS Preview、CSA等文摘数据库。因此,如何实现分布和异构系统间的互操作和信息共享成为图书馆数字资源建设的重点,数据整合正是在这种背景下应运而生的,数据整合的对象是异构数据源,其目的是消除不同数据库系统的分布性和异构性,在逻辑上或物理上进行有机地集中,屏蔽各种数据源的差异,让这些异构系统充分融合,并以统一的查询方式表现出来,达到异构数据的共知和共享。^[3]数据整合一般采用基于数据仓库的整合方式和基于中介模式的整合方式,图书馆数字资源经过数据整合为用户提供了统一简单透明的表示和检索方式,为图书馆异构数字资源的共知与共享问

题找到了很好的解决方案。国内具有影响的异构数据库集成检索系统是CALIS(数字图书馆资源统一检索系统)、国家科学数字图书馆CSDL(跨库集成检索系统)以及台湾“华文知识入口网”,这三个系统都是采用多种集成技术和协议标准构造一个中介结构模式,实现了实体资源分散下的虚拟逻辑数据整合。

2.2 信息整合

数据整合虽然解决了异构数据库中信息实体的合并问题,但这些信息实体间的多种联系如参照关系、引用关系等并没有得到有效的整合。例如,读者通过中国期刊全文数据库找到马费成先生的《论情报学的基本原理及理论体系构建》一文,还想查阅该文的参考文献《情报学若干问题辨析》的原文,然而,该文存在于其它资源系统中。要想解决以上信息关联问题,必须对数字资源进行信息整合。信息整合是指通过某种机制或标准,集成、描述、链接不同类型、不同来源、不同载体的数字资源,使相对独立的信息实体之间产生联系,实现数字资源的全方位整合和一步到位的获取。其整合的对象是信息实体间的关系,如文献的引证关系、文献的责任者与该作者的所有文献与相关文献的关系。信息整合可分为基于信息链接的整合和基于信息门户的整合两种方式。^[4]信息整合让图书馆数字资源信息实体间的复杂关系进行集成和链接问题得到了很好的解决,我国国家科学数字图书馆(CSDL)的资源环境、物理数学、生命科学、化学、图书情报等系列学科信息门户都是信息整合的例子。

2.3 知识整合

信息整合虽然解决了信息实体间复杂关系的整合问题,但其整合的深度不够,没有揭示信息实体间的概念和语义关系,信息整合的结果是信息的集合,而不是知识的集合,不能满足用户的知识需求。要想很好地解决用户的知识需求问题,必须对数字资源进行深层次的知识整合。知识整合是以知识组织方法为指导,以数据整合、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。知识整合的对象是内在于知识体系中的知识结构的观念及关系,目的是实现知识的有效获取,满足人们的知识需求。主题图、本体是新型知识组织体系,例如国内CSDL学科信息门户的发展已经考虑应用主题图技术形成资源集合的主题图。本体由于其具有知识组织体系的功能,能够实现对知识结构的描述与揭示,从而成为知识组织的主流技术。^[5]

数据整合——信息整合——知识整合三个层面的数字资源整合方式是随着对数字资源加工程度的深化和技术方法的进步而概括出来的。基于数据的资源整合是资源整合体系的最低层面,信息整合是资源整合体系的

中间层面,知识整合处于资源整合体系中最高层,是基于数据、基于信息的整合发展到一定阶段的必然要求。这三个层次的资源整合是逐次递进又互有交叉重合的,其发展过程反映了由基于集成服务的数字图书馆向基于面向用户提供知识服务的数字图书馆转型的轨迹。

3 图书馆数字资源整合的模式

根据以上图书馆数字资源整合层次的分析 and 我国图书馆数字资源的实际应用,资源整合的方法可分为以下五种类型。

3.1 基于OPAC的数字资源整合

OPAC(联机公共检索目录)是图书馆数字化建设中的最基础的一项建设。我国大多数图书馆尤其是高校图书馆都建立了OPAC系统,是图书馆众多资源中利用频次最高的,在OPAC中有图书馆所有的馆藏目录,以OPAC系统为基础,整合其他文献资源是提高图书馆数字资源利用率的有效方法。OPAC进行检索时最大的优点是读者感觉不到在进行跨馆际查询,因为只有一个统一的查询界面,非常方便地使用到馆外的或数字化的文献资源,而无须花时间和精力熟悉新的系统和操作方式。如上海交大图书馆的Webpac检索系统,可在统一检索界面上通过著者、题名、主题、关键词、ISBN等检索途径检索上海交大图书馆、华东师范大学图书馆、复旦大学图书馆和上海图书馆的OPAC书目信息,不需要在各个图书馆不同的OPAC界面间来回切换。^[6]基于OPAC的数字资源整合从内容上讲,可分为馆外整合与馆内整合。馆外整合的实质是实现本馆与不同的异构OPAC数据库的整合,当前较多地采用Z39.50协议来完成,聚合不同的OPAC系统,整合生成联合的馆藏书目查询系统,主要用于传统书目查询系统之间的整合。2000年10月,中国国家图书馆采用Z39.50协议,率先在国内图书馆界开始资源共享的研究,我国很多高校图书馆也安装了Z39.50服务器端系统,如清华大学的检索系统可以实现对科学文摘和剑桥科学文摘等进行联合检索,馆内整合是指实现OPAC书目信息与数字资源的整合,主要是在MARC记录里增加856字段——“电子资源地址与存取”字段,记录被著录的数字资源的存取地址和存取方式,实现在实体馆藏中揭示并链接全文电子文献的目的。^[7]

3.2 基于跨库检索的数字资源整合

基于跨库检索的数字资源整合模式通过对不同数据库的不同检索界面、检索方式、检索式构造规则、检索算符、检索字段等进行整合统一检索,大大提高读者

对信息资源获取的效率。跨库整合检索可分为两个层次:第一层次是检索界面整合;第二层次是实现数字资源系统间的分布式异构整合检索。实现跨库整合可借助某些软件,如SFX(Special Effects Cinematography),^①目前美国30多家图书馆使用该软件实现了跨库检索,一批世界著名的信息供应商如ISI、UMI、UBSCO等已开始使用SFX技术和OpenURL协议。全球17个国家近200个机构选择了SFX和MetaLib的解决方案。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链接。它允许用户在数据库中点击一篇文章的记录,然后显示所有能够得到的与这篇文章相关的服务选项列表,该列表可以包括这篇文章在其他数据库的网上全文。^[8]SFX可提供的数字资源包括:全文数据库、文摘、索引、引文数据库、图书馆网上目录系统、其它WEB资源如电子出版系统、网上免费全文等。

3.3 基于资源导航的数字资源整合

基于资源导航的数字资源整合就是通过数字资源的URL建立数字资源的导航系统,将数字资源的检索入口整合在一起,根据数字资源类型不同,建立电子期刊、电子报纸、会议论文集等不同的资源导航库,提供按信息资源名、关键词、资源标识等获取资源的途径。如电子期刊,读者可以按刊名、ISSN号、来源数据库、学科分类等途径浏览检索期刊。当前我国高校图书馆以期刊数字导航系统和数据库导航系统为主,比如北京大学图书馆的西文电子期刊导航系统,读者就可按刊名、ISSN号检索期刊,也可按期刊刊名、学科、出版商/数据库浏览期刊。资源导航系统一般都有以下几个基本功能:字顺浏览功能、分类浏览功能、关键词检索功能,这三个基本功能将帮助读者迅速找到信息资源,并利用超文本链接提供检索入口,对该资源进行全文或目录检索,这种方法解决了资源整合层次中的数据整合问题。

3.4 基于超级链接的数字资源整合

基于超级链接的数字资源整合就是利用网络超文本链接特性,将文献的有关知识点链接起来,达到将有关的信息资源链接在一起的目的,形成一个具有内在联系的有机整体,以方便读者利用各类信息资源。学术论文通过引用与被引用关系形成复杂的引文网络,如果能在信息资源中利用超链接的特性通过参考引文把所有资源都联系起来,形成一种反映各知识点之间直接和间接关系的知识结构性网络体系,对于学术研究将是非常有价值的。理想的引文链接以参考文献为线

①SFX是新的网络电子资源无缝链接整合软件系统,也称上下文敏感参考文献链接解决方案。SFX系统是由比利时根特大学的H.萨姆保尔和他的同事于2000年研发的链接软件。2002年2月,美国EX Libris公司从根特大学取得了SFX链接软件的独占权。

索, 将所有信息资源都整合成一个具有知识关系的网络, 是一种非常理想的、独特的整合方法, 这种方法解决了资源整合层次中的信息整合问题。

3.5 基于知识本体的数字资源整合

要想实现资源整合层次中的知识整合问题, 必须构建基于本体的数字资源整合模式, 本体论(Ontology)是哲学的一个分支, 是一门有关存在及其本质和规律的科学。本体是知识术语的集合, 包括词汇表、语义关系和一些简单的推理和逻辑规则。作为信息资源分类的基础, 本体的目标是捕获相关领域的知识, 提供对该领域知识的共同理解, 确定该领域内共同认可的词汇, 并从不同层次的形式化模式上给出这些词汇和词汇间相互关系的明确定义。基于本体论的知识组织理论能揭示主客观知识及其相互关系, 知识本体即描述概念及概念间的关系, 知识本体的实质是反映概念及其之间关系的知识系统。基于知识本体的数字资源整合模式起的作用: 一是知识本体提供了对领域知识的规范描述和共同理解, 使得知识可以共享, 并可对这些知识进行重用或复用; 二是知识本体使得不同领域的知识体系化、结构化、形式化, 为实现数字图书馆数字资源整体化、一体化整合奠定了基础。^[9]

本体的引入使得自动化加工处理从传统的语法层次上升到语义层次。通过各知识库的基于本体知识关联和逻辑推理等知识链接手段形成知识网络。在人与软件代理之间、人与人之间建立知识链接, 使本领域的知识得到共享和重用, 使领域内不同的系统、模型之间能够进行互操作, 为各领域或领域之间通过概念体系提供一种语义路径图, 实现数字资源基于语义的整合和利用。

如何构建知识本体, 比较有效的方法是以现有本体为基础构造出新本体, 分类表、词表等传统的知识体系就是简单的知识本体, 元数据方案也是本体的简单表现形式, 根据不同类型的资源制定相应的元数据方案这一过程就是知识本体的构建过程。^[10]

4 图书馆数字资源整合模式的比较

由于数字资源的内在联系不是单向或线形的, 而是呈网状的复杂关系, 因此不可能通过单一的整合方法来描述数字资源的全部, 必须通过多种整合方法来多角度和多层次地挖掘和揭示这些内在关系。而对现有的图书馆数字资源整合模式进行分析比较, 有利于了解各种整合模式的特点, 指导本单位图书馆的数字资源整合工作, 进而对图书馆数字资源整合模式进行合理优化。

以上整合方式各有优缺点。基于OPAC系统的整合充分利用了OPAC系统高访问率的优势, 能实现在馆藏

书目数据中反映数字资源信息, 特别是对全文电子期刊全文的链接, 可以提高数字资源的利用率。但这种检索模式检索的层次不高, OPAC系统一般遵循的Z39.50协议尽管从理论上能实现全文检索, 但目前在图书馆界的应用一般以目录层次为主, 这注定它直接可实现的整合检索以书目为主, 如把数字期刊整合进OPAC系统一般都只提供该刊的数据库检索入口, 而不能直接提供其中某篇文献的全文链接。^[11]

基于跨库检索系统的整合模式能实现在多个数据库内统一检索, 可以为不同资源访问提供统一的检索入口, 读者可以选择自己感兴趣的检索目标, 提交统一的检索请求, 实现多个数据库的同时检索, 一次认证、一次检索、得到全部检索结果是基于跨库检索整合模式的优点, 但这种整合系统受各类数据库系统搜索引擎的限制, 而且分布式检索系统是跨库整合检索系统的理想模式, 需要广泛范围内的协作和遵循必要的元数据标准及互操作协议, 实现起来还有困难。^[12]

基于资源导航的整合模式是较低层次的整合模式, 虽然可以解决图书馆数字资源数据整合的问题, 但它只是为对于资源不够了解的读者提供检索入口, 为进入各个信息资源提供了方便, 在具体的检索过程中, 读者还是要进入各个资源进行重复性的检索操作和筛选结果, 并没有真正的将分布的资源整合在一起, 不能真正实现“一步到位”的检索。^[13]

基于链接系统的整合模式的优势是能从参考文献的角度出发整合图书馆的信息资源, 为研究者提供很多便利, 但用文后参考文献的链接查找资料, 容易造成漏检; 另外读者利用链接进行资源检索时, 由于链接层次的不断深入, 容易迷失方向, 浪费读者的时间。

基于本体的数字资源整合一种基于语义的信息集成。是数字资源组织的理想境界, 未来的基于本体的知识网格为用户提供基于语义的服务, 大大提高了信息加工和服务的知识含量, 为知识创新提供了坚实的基础。

通过对以上五种整合模式的比较, 结合数字资源整合体系的目标层次分析, 对于图书馆数字资源整合未来的发展, 知识整合将是数字资源整合的主流趋势, 语义网作为以知识资源为基础的人类知识整合的大环境, 是人类知识的一个宏观网络, 而本体又是语义网构建的核心和基础, 所以领域本体的构建、基于本体的数字资源整合将成为知识整合的研究重心。

5 研究趋势展望

通过对以上我国图书馆数字资源整合研究来看, 数字资源整合层次分析以及对整合模式发展, 其实是图书馆信息服务理念由“以图书馆 (下转第82页)

- [10]谷歌瞄上微博客实时搜索领域[EB/OL].[2009-07-15].<http://news.ctocio.com.cn/421/8909421.shtml>.
- [11]微博客“Twitter”成英语词典新宠[EB/OL].[2009-07-15].<http://news.hsw.cn/system/2009/07/07/050233073.shtml>.
- [12]美前国家安全顾问:Twitter应获诺贝尔和平奖[EB/OL].[2009-07-15].<http://www.17tech.com/news/20090712122151.shtml>.
- [13]美宇航员将在太空发布Twitter信息[EB/OL].[2009-07-15].<http://tech.sina.com.cn/i/2009-07-03/11023233214.shtml>.
- [14]IT频道. 美宇航员从太空向地球发送首条Twitter信息[EB/OL].[2009-07-15].<http://it.rednet.cn/a/8851357.shtml>.
- [15]微博客twitter风靡海外 教你如何玩twitter[EB/OL].[2009-07-15].http://news.xinhuanet.com/internet/2009-06/25/content_11599748_1.htm.
- [16]微博客: 成长的烦恼[N].计算机世界报, 2009-04-20(10).
- [17]Twitter旋风吹到中国 微博客: 下一个网络新贵?[EB/OL].[2009-07-15]. [EB/OL].[2009-07-15].<http://focus.scol.com.cn/zgsz/20090712/2009712144930.htm>.
- [18]the shifted brarian [EB/OL].[2009-07-15].<http://www.theshiftedlibrarian.com/>.
- 作者简介: 李华(1964-),女,石家庄学院图书馆馆员;赵文伟(1958-),男,河北省科学技术情报研究院研究员。

(上接第69页) 为中心”向“以用户为中心”转变的过程。作为互联网发展的新阶段,Web3.0的概念已被提出,如果说,Web1.0以门户网站为中心,Web2.0以用户为中心,那么,Web3.0的精髓就是以服务为中心。^[14] Web3.0将为图书馆提供更多可能的平台,图书馆数字资源整合随着Web3.0信息环境的变化而变化,在取得一定进展的同时,图书馆数字资源在整合过程中主要障碍在于不能对数字资源的语义进行描述,如果能有效地构建一种带语义的服务环境,实现数字信息资源处理的智能化,则可以产生一种更能符合用户需求、更准确的数字资源整合模式,而语义网的应用就可以实现这一目标。^[15]语义网的核心是元数据,它通过在现有网络基础上增加共用的、标准的机器可理解的元数据,使得原来网络环境下难以实现的许多应用成为可能或更为有效,如信息获取、信息过滤、网络自动服务等,计算机可以对所搜集到的数字资源进行智能化的评估,可以像人脑一样“理解”数字资源信息的含义,完成“智能代理”的功能,从而数字资源整合的结果更能符合人的实际需求。因此,Web3.0环境下基于语义网的数字资源整合将是未来的研究发展趋势。目前,发展语义网的两种主要技术——XML和RDF已经产生,但更主要的技术难题在于要让计算机进行更准确的“思考”和“推断”,语义网在数字资源整合中的应用还需面临语义网格、SOA、本体等关键技术的挑战。^[16]

参考文献:

- [1]黄晨.资源整合模式及其实现研究[J].大学图书馆学报,2004(1):25-28.
- [2]王知津.从情报组织到知识组织[J].情报学报,1998,(6):231-234.
- [3]马文峰,杜小勇.基于数据的资源整合[J].情报资料工作,2007(1):41-45.
- [4]马文峰,杜小勇等.基于信息的资源整合[J].情报资料工作,2007(1):46-50.
- [5]马文峰,杜小勇等.基于知识的资源整合[J].情报资料工作,2007(1):51-56.
- [6]于凤英,王召龙.浅谈图书馆数字资源整合[J].济南职业学院学报,2005(6):70-72.
- [7]陆永兵.高校图书馆信息资源整合研究[D].吉林:东北师范大学硕士论文,2006.
- [8]张健兰,张新宇.图书馆数字信息资源整合初探[J].晋图学刊,2004(4):36-38.
- [9]陈红梅.基于本体的数字资源整合研究[J].情报杂志,2007(11):151-153,156.
- [10]羊照生.图书馆数字资源整合研究[J].科技情报开发与经济,2008(1):1-3.
- [11]朱萍.馆藏数字资源整合方案研究[D].郑州:郑州大学硕士论文,2005.
- [12]李阳晖.面向用户的图书馆信息资源整合模式分析[J].情报方法,2005(10):34-35.
- [13]刘峥.数字资源整合的现状及其发展[J].图书情报知识,2003(10):40-41.
- [14]Xi Shi, Patricia J. Holahan. Satisfaction Formation Process in Library Users: Understanding Multisources Effects [J]. The Journal of Academic Librarianship, 2004(2):22-26.
- [15]Zhuge Hai.China's e-science knowledge grid environment[J].IEEE Intelligent Systems, 2004,19(1):13-17.
- [16]王宁.语义网的研究与展望[J].科技情报开发与经济,2007(32):1-3.
- 作者简介: 黄传慧(1974-),女,武汉大学信息管理学院2006级博士生,湖北工业大学管理学院讲师;李娟(1976-),女,硕士,清华大学图书馆馆员。