

# 文献资源集成揭示的模式与应用

宋文, 傅红梅

(中国科学院国家科学图书馆, 北京 100080)

**【摘要】**通过分析文献资源集成的原理和目前主要的集成模式, 总结出集成揭示体系的建设内容: 文献资源集成组织体系的建设、文献资源集成加工体系的建设 and 文献资源集成服务体系的建设, 阐述了中科院国家科学图书馆集成揭示系统的实施策略。

**【关键词】**文献资源; 信息资源; 集成揭示

**【中图分类号】**C250.7 **【文献标识码】**A **【文章编号】**1002-1167(2008)01-0073-04

文献信息资源集成揭示是计算机技术和数字图书馆发展到一定阶段的必然结果。通过无缝整合数字图书馆的服务和广泛的信息资源, 为用户提供充分方便的集成检索, 无缝链接, 动态扩展和联合服务, 是集成揭示体系建设的目的。

## 1 相关研究文献综述

笔者在 LISA 数据库中用 (Integrat \* or Aggregat \*) 进行检索, 在用年代作限定后, 分别得到如下结果: 从 LISA 最早收录年代到 1969 年共检索到 25 篇文献; 70 年代共有 128 篇文献; 80 年代有 386 篇文献; 90 年代上升到 826 篇文献; 从 2000 年到 2006 年已经达到 758 篇文献。笔者大致对 2000 年到 2006 年的 758 篇文献进行浏览, 发现涉及的内容非常广泛, 包括企业信息集成、数据库集成、系统集成、网络资源集成、各专业领域信息集成等等。可以看出集成已经成为数字图书馆建设的主旋律。

笔者用 (Aggregat \* or Integrat \*) 与 Digital library、Documentation、Resources 进行组配检索, 共检索到 51 篇与文献资源集成揭示相关的文献。主要讨论的问题有: 纸本资源和电子资源集成揭示、电子资源的编目、各种类型电子资源的集成揭示和服务、电子资源管理系统、各类型资源的元数据集成等问题。例如: Janice Lewis Steed 和 Clark Nall 的论文探讨在图书馆构建一个主题指南数据库, 包含所有格式的资源, 作为用户查找资源的入口, 可以检索到图书馆的所有的资源。<sup>[1]</sup>Davies, S 和 Wales, A 在论文中详细介绍了英国 NHS 图书馆联盟联合的电子资源建设和集成虚拟图书馆建设工作。<sup>[2]</sup>

笔者在维普中文科技期刊数据库中, 共检索到 200 余篇与集成相关的论文, 主要讨论的是关于网络信息资源整合、电子资源整合的技术、模式和方法等问题。理论和方法探讨的论文比较多, 实践研究的论文相对较少。

## 2 文献资源集成揭示模式分析

文献资源的集成从原理上可以分为三种类型:

(1) 物理集成: 将资源的内容物理上整合到一起。比如全文数据库, 有些论文中也将元数据仓库认为是物理集成。

(2) 逻辑集成: 数据是分散存储的, 将不同来源的数据通过逻辑对应关系或链接关系逻辑上集成为一个有机整体。如跨库检索系统就是典型的逻辑集成方式, Web39.50、CrossRef、OpenURL、唯一标识符机制、超文本链接、通过知识组织体系技术等都是实现的信息资源的逻辑集成。

(3) 物理集成和逻辑集成结合: 元数据仓库是典型的物理集成和逻辑集成的结合, 还有 WebService 模式的集成、搜索引擎、馆藏总目录、联合目录、知识库等。

目前文献信息资源主要的集成模式有:

### 2.1 集成目录

集成目录是集成各种类型文献的一种有效方法。目前国内外大多数图书馆都将电子资源、纸本资源集成到一个目录体系下。OCLC 在研究 FRBR 结构在目录组织中的应用,<sup>[3]</sup>国外一些主要的图书馆集成系统也在尝试实现 FRBR 化的记录显示, 如 Innovative 公司的 Millennium 系统。<sup>[4]</sup>中科院国家科学图书馆建立的集成期刊目录就是将纸本期刊和电子期刊集成到一个目录体系下提供服务。

### 2.2 知识库和 OpenURL

知识库存储了资源对象的元数据信息、资源对象的关系、链接规则、服务规则和用户推理的相关信息, OpenURL 是不同系统之间传递元数据信息的一种标准。知识库和 OpenURL 结合, 通过服务解析器, 实现来源资源与目标资源的、针对特定用户的情景敏感的连接。<sup>[5][6]</sup>

### 2.3 唯一标识符和 CrossRef

唯一标识符是因特网上用来唯一标识对象的机制, 通过唯一标识符机制, 可以实现因特网数字对象之间的互操作。<sup>[7]</sup>DOI 是美国出版协会建立的用于标识数字对象的唯一

标识符系统, CrossRef<sup>[8]</sup>是商业性的 DOI 登记注册系统。CrossRef 系统与本地的知识库和解析服务系统结合, 通过 DOI 和 OpenURL 传递数字对象元数据信息, 达到资源之间更加广泛、顺畅和高效的互连和集成。

#### 2.4 元数据仓库

元数据仓库是将各种资源对象元数据集中到一个仓储系统, 提供用户对所有资源的集成检索, 通过检索结果引用到分布式的资源对象。这种资源集成服务方式可以大大提高资源的检索效率和资源被用户发现程度。

#### 2.5 跨库检索系统

跨库检索系统实现集成的方式与元数据仓储方式相反, 所有数据都分布在不同的应用系统中, 在用户提交检索请求后, 系统自动向各应用系统提交检索请求, 再将结果集中进行查重、合并和排序, 以统一的界面提交用户。这种方式的缺点是系统响应速度慢。

#### 2.6 知识组织体系

通过知识组织体系, 从知识概念的关系出发实现资源的集成揭示是当前数字图书馆的又一个研究热点。我们在中科院文献情报中心前瞻性研究课题“知识组织体系技术及其应用机制”研究中作了一些实践性探索, 建立了集成知识组织体系和知识导航系统, 对分布式资源进行学科体系导航。

#### 2.7 WebService 模式

WebService<sup>[9]</sup>是基于万维网的分布式计算和服务集成技术。数字图书馆可以将 WebService 标准和技术用于文献信息资源的集成服务。在这样一个技术体系下, 各种资源和资源服务系统可以是分布式的, 其核心是建立基于 UDDI 标准的注册登记系统, 提供资源和服务的发现机制和集成服务机制。目前 WebService 技术在图书情报领域的应用还处于试验阶段, 真正推广使用需要各种应用系统能够提供基于 WebService 标准的服务。

以上我们讨论了文献信息资源集成的多种模式, 实际设计文献资源集成揭示系统时, 并不是选择一种集成模式, 大多数情况下是多种集成模式同时使用, 各种模式相互补充发挥作用。比如跨库检索系统与 WebService 可以结合, 跨库检索系统、WebService 系统又可以与知识组织体系结合使用。知识库与集成目录可以相结合, 知识库技术与 CrossRef 和 DOI 结合等等。究竟如何设计集成揭示系统, 需要根据资源状况, 需求状况和技术应用水平, 甚至是时间、人力、经费等因素进行综合考虑。

### 3 对文献资源集成揭示体系的思考

文献信息资源的出版和发行模式正在发生巨大的转变, 各种载体资源并行出版发行, 出版商不仅提供信息内容, 同时提供集成服务平台, 文献信息资源的获取不再仅仅依靠馆藏资源等等。这些变化使得文献资源集成揭示体系建

设已经成为迫在眉睫的任务。文献资源集成揭示体系要体现信息资源出版和发行模式转变带来的服务模式的转变。通过集成、整合、链接, 大大提高已有的服务能力。

文献资源集成揭示体系的建设主要包括三个方面: 文献资源集成组织体系的建设、文献资源集成加工体系的建设 and 文献资源集成服务体系的建设。集成组织体系是整个集成揭示体系的基础; 加工体系是实现文献资源集成揭示的保障; 集成服务体系是文献资源集成揭示体系建设的最终目标。

#### 3.1 文献资源集成组织体系的建设

文献资源的集成揭示要实现多载体、多格式、多类型资源的集成揭示 (如电子刊与纸本刊); 资源来源渠道的集成揭示 (如同一种电子刊来源于不同的全文数据库); 物理馆藏与虚拟馆藏的集成揭示 (物理馆藏一般只需要对应到收藏单位, 但虚拟馆藏需要对应到资源来源渠道和用户, 甚至是某个具体的 IP 地址); 不同层次资源的集成揭示 (如书目层、目次层、全文层); 资源与服务的集成揭示 (资源和服务是互补的, 服务可以转化为资源) 等。在这样一个集成揭示目标下, 需要解决的资源组织体系方面的问题有:

##### 3.1.1 文献信息资源的元数据描述体系

不同类型、不同格式、不同载体、不同层次资源对象的元数据描述体系建设。在元数据描述体系中, 不仅要确定各种类型资源的形式描述体系 (如题名、作者等的描述), 更重要的是要对各种资源对象的相互关系进行描述和揭示。通过对资源和资源关系的全方位揭示, 使用户可以从一个资源引用到各种相关资源。

##### 3.1.2 唯一标识符应用机制建设

在文献信息资源的元数据描述体系中揭示了资源的逻辑关系, 但要实现计算机系统对各种资源的自动链接和关联, 需要实现各种资源对象的唯一标识符体系建设。国际上目前有许多唯一标识符体系, 如 SICI, BICI, DOI, ISSN, ISBN 等, 我们重点要解决的是唯一标识符体系的选择和本地化应用问题。

##### 3.1.3 资源知识内容的组织体系

在对资源形式和关系进行揭示后, 还需要从学科体系或主题体系, 也就是知识关系角度, 对资源和资源的关系进行揭示。这就需要解决知识组织体系建设和应用的问题。

##### 3.1.4 资源描述体系的规范化体系

规范体系用于资源描述中的描述项目和元素的规范。规范体系一般包括个人名称规范、机构名称规范、资源类型规范和资源格式规范等。

资源描述体系、知识组织体系和规范体系是资源集成组织的三个重要方面。

#### 3.2 文献资源集成加工体系的建设

集成的文献组织体系需要依靠集成的文献加工体系去



步检索, 获得结果集; 第2步, 点击目标资源或目标服务获取资源或服务。如果需要扩展链接可以增加步骤3。

(3) 每当用户提交一个检索式, 或进行一次链接点击后, 需要后台多个资源库的联合支持和系统的多个步骤来实现。

(4) 在用户提交检索式后, 系统根据用户提交的检索内容, 自动到联合目录库、知识库、规范数据库、元数据库中检索, 然后到知识库中进行情景推理, 将检索结果和推理结果按用户获取资源的方便程度排序。

(5) 当用户需要实现从一种资源到另一种资源的链接时, 系统又会根据各种资源的链接关系, 根据底层的知识组织体系、规范库等检索相关资源, 接着又从知识库得到用户最合适的资源和服务, 提供给用户。

(6) 当用户最终需要获取资源和服务时, 系统又回到知识库, 提取该资源或服务的链接规则, 链接到用户需要的资源或服务。

[参考文献]

[1] Lewis, Janice Steed; Nall, Clark, Integrating print and electronic resources: Joyner Library's PirateSource. Acquisitions Librarian, 2005,

16 (33/34): 107 - 120

[2] Davies, S; Wales, A. Integrating electronic information resources for NHS Glasgow staff at the point of need: a model of interlibrary collaboration and resourcesharing. Health Information and Libraries Journal, 2001, 18 (4): 229 - 37

[3] OCLC Research Activities and IFLA's Functional Requirements for Bibliographic Records. [2007 - 01 - 04]. <http://www.oclc.org/research/projects/frbr/default.htm>

[4] Millennium 2005 Preview: FRBR Support. INN - Touch, 2004, 18 (1): 9

[5] SFX context sensitive linking. [2006 - 08 - 04]. <http://www.exlibrisgroup.com/sfx.htm>

[6] SerialSolutions. [2006 - 08 - 04]. <http://www.serialsolutions.com/home.asp>

[7] 毛军. 数字资源唯一标识符的现状与发展. [2006 - 08 - 05]. <http://cdls.nsl.gov.cn/cdls2/w3c/2003/UI/>

[8] CrossRef.org. [2006 - 08 - 05]. <http://mailserv.crossref.org/index.html>

[9] W3C. Web Service Activity. [2006 - 08 - 05]. <http://www.w3.org/2002/ws/>

Research and Practice of Document Resources Integration

SONG Wen, FU Hong - mei

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

**Abstract:** The paper analyzes main integration modes at present and put forward integrated system should chiefly includes the document resources integrated organization system, document resources integrated processing system and document information resources integrated service system. In the end, the implementary strategies of the CAS's document information integrated system are illustrated.

**Keywords:** document resources; integrated organization

[作者简介] 宋文 (1961 - ), 女, 研究馆员, 中国科学院国家科学图书馆资源建设部副主任。

[收稿日期] 2007 - 06 - 26

(上接第69页)

源进行管理, 使高校图书馆解决读者占座问题找到了突破口, 采用现代管理方式不但能解决占座现象, 而且对建设和谐图书馆, 对文明读者的培养造就良好的软环境具有积极意义。图书馆座位资源利用计算机管理, 还需要图书馆界共同关注、共同努力, 也需要社会力量的广泛支持。

[参考文献]

[1] 贾成娟, 韦成礼. 藏借阅一体化管理模式下的抢占座位现象

[1]. 农业图书情报学刊, 2005, (5).

[2] [3] 朱希伟. 图书室的免费“午餐”——占座现象透视 [J]. 浙江统计, 1999, (8).

[4] 袁闯, 王月娥. 关于大学生占座状况的调查——以吉首大学图书馆调查为例 [J]. 农业图书情报学刊, 2006, (5).

[5] 刘大杰, 刘金鹏, 孙小菡. 基于单片机和红外接近开关的图书馆座位管理系统 [J]. 电子工程师, 2006, (5).

Probing of Computerized Management of the Seats in Library

QU Kui, LI Chun - li

(Library of Dalian Polytechnic University, Dalian 116034, China)

**Abstract:** All the resources in a library, including documents, categories, cadres and work result, are digitally managed, except seats. Seats in libraries are mostly used unsystematically. Computerized management of seats in a library is an effective way to resolve the problem of seats occupying.

**Keywords:** library; computerized management; seats occupying; seats resources

[作者简介] 曲奎 (1961 - ), 男, 现在大连工业大学图书馆工作, 已发表论文 10 篇。

[收稿日期] 2007 - 07 - 02