

● 黄国彬<sup>1</sup>, 孙 坦<sup>2</sup>

(1. 中国科学院 文献情报中心, 北京 100080; 2. 中国科学院 研究生院, 北京 100080)

# 建设复合图书馆的技术支撑体系

【关键词】复合图书馆; 技术支撑体系; 电子图书馆; AGORA 项目

【摘要】紧紧围绕“基本功能决定图书馆所需的技术”这一主线, 从技术角度入手, 集中研讨英国于“电子图书馆”第三阶段所启动的五个复合图书馆原型项目之一——AGORA 项目, 以求得出对我国今后开展复合图书馆项目大有裨益的经验与规律性认识, 并尝试构建一种符合我国图书馆发展事业实际情况的复合图书馆技术支撑体系。

【中图分类号】G250

【文献标识码】A

【文章编号】1005-8214(2005)06-0008-03

## 1 概述

### 1.1 复合图书馆技术支撑体系概述

图书馆事业的发展历程表明, 图书馆的资源组织、服务提供、人员培训与管理等各项业务的顺利开展都离不开技术的应用。正是因为与计算机技术、通信技术以及其它各项先进技术的相互融合, 并不断地对这些技术加以吸纳应用, 图书馆才得以提高服务水平。<sup>[1]</sup>

通常人们所说的技术体系, 即是指由一系列技术有机组合而成的、具有系统性的技术整体。对于复合图书馆来说, 它的技术支撑体系就是保证它的功能得以充分发挥的技术整体。其实质是复合图书馆在网络和计算机上的具体实现, 是传统图书馆与数字图书馆共同存在时, 处理、存储和应用传统印本资源与数字化资源的基本框架。它是建设复合图书馆的基础。

### 1.2 基本功能决定复合图书馆的技术支撑体系

资源和服务就是数字图书馆的基本功能。<sup>[2]</sup>到目前为止, 学界也认为, 复合图书馆的“亮点”主要体现在两大方面, 一是无缝存取 (seamless), 二是综合集成 (integrated)。而蕴含于这两者之中的实质含义则是: 复合图书馆主要是整合传统印本资源与数字化资源 (强调的是资源), 以综合集成的方式给本地和远程的用户提供信息和知识服务 (强调的是服务)。尽管人们对复合图书馆的“亮点”表述得较为抽象委婉, 但万变不离其宗, 复合图书馆与数字图书馆在基本功能上是相似的, 即它最终仍须紧紧地抓住两点——资源与服务。

我们认为, 复合图书馆作为一种图书馆发展形态, 在建设的过程中, 会涉及需要运用技术才能妥善解决的信息

采集、加工、处理、存储、组织、发布、利用 (含版权保护) 和归档等问题。在解决这些问题的过程中, 建设一个科学可行的技术支撑体系显得极其重要。

当然, 必须认识到, 尽管复合图书馆是对传统图书馆与数字图书馆的有机糅合,<sup>[3]</sup>但绝不是这两者的简单叠加! 恰恰相反, 它不仅海纳了两家之精华, 而且也拥有自身的特色。作为对这些特色的继承与维持, 复合图书馆的技术支撑体系不仅保留有传统图书馆与数字图书馆技术支撑体系的痕迹, 更有着自身的特色。

### 1.3 有关复合图书馆技术支撑体系的研究现状

从当前的研究情况分析, 自台湾学者顾敏先生将复合图书馆介绍到国内起, 以初景利教授、黄宗忠先生为代表的一大批国内学者纷纷撰文, 抒写自身对复合图书馆的认识与见解, 并由此掀起了一股继数字图书馆研究热之后的“复合图书馆”研究热。笔者在撰写本文的过程中, 采用任意字段检索的方式, 输入“复合图书馆”这一词组, 在维普数据库中竟然检索到大约 110 篇有关复合图书馆的文章。然而, 很遗憾的是, 迄今为止, 围绕建设复合图书馆所需的技术而展开讨论的文章却是少之而少, 甚至可以说之为零。

## 2 复合图书馆面临的技术问题

不管是技术, 抑或是理论, 其本质都是人们用来解决问题的工具。发展复合图书馆, 理所当然地, 我们会遇到相应的问题。而这些问题出现, 归根到底是由复合图书馆的本质、资源与服务所决定的。人们正是在不断运用技术应对、解决这些问题的过程中, 逐步建立技术体系, 并不断完善的。那么, 发展复合图书馆, 究竟有哪些问题需要由技术来解决呢?

我们认为, 复合图书馆既然是传统图书馆与数字图书馆的有机结合, 那么, 它所面临的、需要由技术来解决的问题自然而然地包括了以下三大类: 为了维持自身的服务特色与功能而遇到的问题; 为了传承传统图书馆的特色而遇到的与传统图书馆所面临的问题相类似的问题; 为了弘扬数字图书馆的特色而遇到的与数字图书馆所面临的问题相类似的问题。

综合目前各方面的情况分析可知, 复合图书馆需要解决的两个核心问题就是互操作与整合的问题。即在复合图书馆中, 如何实现对分布异构的数字化信息资源进行集成检索和整合浏览, 并在“一站式”检索的基础上支持用户对分布异构的多层次、多类型、多格式的数字信息以及信息服务进行个性化定制。因此, 基于各种元数据的开放描

【基金项目】本文为国家社科基金《复合图书馆理论与我国复合图书馆建设模式研究》(03BT0005) 研究成果

述、开放链接、互操作等方面的技术问题则是发展复合图书馆必须面对的。

### 2.1 复合图书馆的互操作问题

早在2000年,笔者就曾撰文指出,在数字图书馆的研究方面,今后需要加强的,乃是数字图书馆互操作问题的设计与实现。第二代数字图书馆的发展状况已足以证明这一预测的正确性。时至今日,当人们在谈论如何构建复合图书馆的技术体系时,互操作问题依然是人们无法回避的。即使是诸如英国这样一个在复合图书馆建设方面已投入大量人力并已取得一定成效的国家,在它开展的复合图书馆原型项目——HyLiFe中,也充分体现对互操作问题的重视。<sup>[4]</sup>

有关复合图书馆的互操作问题,Stephen Pinfield等人也撰文指出:<sup>[5]</sup>到目前为止,依然缺乏一种统一的方法来提供和管理这种针对复合图书馆资源的整合访问。用户往往被迫单独同各种服务接触,而且往往由于在不同检索系统中重复相同的步骤而浪费了大量的时间与精力。同时,使用不同的界面也增加了低效率的风险。例如,由于不熟悉某种服务的特性,用户往往不能发现相关的资源。因此,要发展一个完善的复合图书馆,必须妥善解决这一问题。

### 2.2 复合图书馆的整合问题

上个世纪一些单位在建设数字图书馆的过程中,已建设了一些具有一定水准的数据库,将这些数据库无缝地整合到复合图书馆系统中,必将会极大拓展复合图书馆的资源阵营。当然,实现这样的整合,必须解决好各种数据库的异构性、分析性以及复杂性的问题。<sup>[6]</sup>

### 2.3 复合图书馆面临的其它问题

(1) 资源的有效存储。<sup>[7]</sup>由于复合图书馆所面临的数据是海量的,因此,解决庞大的数据存储需要有数据库操作技术的支持。同时,也要有压缩技术的支撑,以提高数据库的存储能力。

(2) 资源的科学分类、索引、检索。复合图书馆所面临的数据类型各不相同。对不同的内容,需要建立不同的分类体系和索引机制。并在此基础上,借助先进的中文搜索、图像搜索、语音搜索、智能搜索等大量的人工智能技术,建立一个高效的检索系统。

(3) 资源的传递与保护。面对用户的服务请求,复合图书馆的服务系统需要在最短的时间作出响应。这需要有诸如网络传输技术、多媒体解压分层传输技术等传递技术的支撑。在传递资源的过程中,还要有诸如数字水印等技术来保护受版权保护的资源。

(4) 开发工具的选择与平台的搭建。这其中会涉及到软件技术、信息输入技术、搜索引擎的技术、知识发现与数据挖掘技术等问题。也只有充分运用好这些技术,作为复合图书馆的“服务门面”——交互式用户界面的设计才能得到成功实现。

此外,需要由技术解决的问题还有,分布式服务与系统的集成与协调,用户认证、收费、付款等问题;技术使用上的统一协议、规范、标准等。随着复合图书馆实践的不断深入,需要由技术解决的新问题也会不断出现,这是一个动态的发展过程。

## 3 可以借鉴的相关技术支撑体系

就目前来看,复合图书馆所面临的、需要由技术解决的主要问题乃是互操作与整合。要解决异构信息之间的互操作,必须建立包括各种硬件、操作系统、应用系统和元数据等不同层面的互操作协议和信息代理机制。同时,面对其它技术问题,复合图书馆完全可以借鉴已有的建设经验予以解决。

### 3.1 第二代数字图书馆的技术支撑体系

由于复合图书馆是对传统图书馆与数字图书馆的有机融合。因此,就复合图书馆所要秉承的数字图书馆的特色与功能而言,第二代数字图书馆所采用的关键技术对它依然有借鉴意义。

笔者在博士学位论文中也就数字图书馆的技术体系问题开展过研究,<sup>[8]</sup>并得出了以下认识:“总体来说,开发数字图书馆需要如下的基本技术:内容处理技术,提供主要信息(原始数字信息和辅助信息(传统非数字媒体)的数字化转换)的创建、存储和检索;信息访问技术,提供对各种信息资源的超越时空限制的有效访问;用户友好和智能接口技术;互操作技术和开放系统开发等技术。”

有学者认为,第二代数字图书馆的技术支撑体系应该是这样的:<sup>[9]</sup>

- 用于信息采集和移值的技术。包括印本信息采集和数字化信息采集两个方面。传统信息采集使用的技术手段包括有:扫描、缩微、光学字符识别(OCR)、语音识别和人工智能等数字化信息的采集包括文本、图像、音频和视频信息的处理。

- 用于信息组织的技术。包括信息组织技术、压缩技术、海量信息存储技术。信息的组织以图书馆为代表,主要应用研究了MARC数据、元数据、XML等技术;信息的压缩涉及文本、视频、音频等多种形态和标准;海量信息存储应用了分布结构。

- 用于信息检索的技术。以全文检索为主。现在,由于人们对语义网研究的逐渐深入,概念检索也日渐成熟,这方面也会涉及到互联网的搜索引擎技术。

- 用于确保信息安全的技术。信息安全涉及到网络安全、信息加密和数字水印等。

### 3.2 英国 e-Lib 第三阶段复合图书馆原型项目 AGORA 的技术支撑体系

AGORA 原型项目由英国东英吉利大学主持,是一个由公司和东英吉利大学以外的5所大学合作开发的一个复合式图书馆管理系统,提供对各种资源和服务的集成存取,包括对复合图书馆载体的跨库检索等。

在AGORA项目的评估报告中,项目组指出,复合图书馆遇到的问题主要有:<sup>[10]</sup>

- (1) 互操作问题。主要是因为各个图书馆对Z39.50的不同使用而引发出来的问题。这一问题又可分为两种情况,一方面,系统提供商对Z39.50标准的使用并不统一;另一方面,由于不同系统提供商对Z39.50标准的不同使用,从而产生出不同的内容标准,以致使解释、排架以及结果集的别重工作困难重重。

- (2) 馆藏揭示的问题。

- (3) 术语学的问题以及缺少由协议控制的词汇表。

- (4) 多年以来悬而未决的授权问题。

- (5) 用户界面的设计。当然,界面的设计有时并不仅仅

仅是技术的问题。但是,必须认识到的是,用户对任意一个系统的接受与否在很大程度上受界面设计风格的影响;界面设计是一个需要加以重视的问题。

考虑到上述问题,该项目将目标定位为:在一个开放式、标准化的平台上研究分布式、混合介质信息的管理问题。AGORA 正逐步开发出一套基于代理模式的复合图书馆管理系统(HLMS)。该系统采用基于 Models Information Architecture 的三层架构,即客户、Web化和智能代理。同时,它所搭建的 Broker 架构,也将 HLMS 作为核心部分予以纳入。

通过 Broker 系统,复合图书馆便能够以一种集成的方式向用户提供异构信息的发现、收藏地点的定位、借阅请求的提示、文献传递等多项服务。具体地说,通过对 AGORA 项目目前已经推出的演示系统的功能的研究,我们认为:复合图书馆的技术支撑体系必须包括以下诸项技术:

- 精湛的 Web 技术。这一技术可谓是 AGORA 项目的“耀眼点”。通过精心设计的 Web 页,用户可以进行检索、提出借阅请求,或获得电子方式以及传统方式的文献传递服务。

- 人性化的检索技术。在检索过程中,用户可以查看并跟踪检索状态;用户可以浏览所有检中记录,并根据需要对之加以排序、分类、复制、聚类。

- 独到的聚类技术。AGORA 提出“信息全图”的概念,意即根据用户的兴趣和需求对资源实行个性化的分组集合、并列成资源图,供用户点击检索。这将大大提高检索效率。

- 先进的资源传递技术。通过检索界面,用户可以根据检中记录直接向服务商提出文献传递请求。利用 URL 可以同时确定和递送可用 Web 检索的电子资源。

由此,我们可以得出以下结论:在复合图书馆的技术支撑体系的建设上,一定不能忽视四大方面的技术,即资源采集与组织的技术、资源检索的技术、资源传递的技术与安全技术。

#### 4 建设复合图书馆的技术体系

复合图书馆将同时提供传统印本资源和数字化资源的服务与利用。<sup>[1]</sup>为此,复合图书馆需要利用各种信息技术对分散化、多样性的信息资源进行有效整合和集成,使多种类型的资源形成一个有机整体,并最终给用户提供一个“一站式”服务。

在提供“一站式”服务的过程中,复合图书馆将传承传统图书馆的功能特色,同时,对数字图书馆的功能优势予以容纳,在此基础上,复合图书馆还有自身的独到之处。因此,在建设复合图书馆的技术支持体系方面,我们尝试从三个不同角度加以论述。当然,这只是一个抛砖引玉的过程,具体应该搭建什么样的技术支撑体系,还需要图书馆界同仁在实践中不断摸索与完善。

##### 4.1 从传统图书馆功能体系的角度出发

依托传统图书馆构建起来的复合图书馆,在开展自身业务的过程中,仍须构建起诸如下列的子系统:如采访、编目、流通、检索、期刊、数据交换、管理信息、联合编目等系统。

由于这些系统都需要一定的技术支撑,而且由于这些子系统功能各异,因此,在使用的技术方面,它们也各不

相同。可以说,复合图书馆的技术支撑体系是渗透在这些子系统当中的。明了发挥各个子系统的功能所需的技术,也就明了复合图书馆的技术支撑体系。对此,我们认为,复合图书馆的技术支撑体系应该包括:用于资源组织加工的技术;用于资源存储的技术;用于资源检索的技术;用于资源传递的技术及用于确保资源安全的技术。

##### 4.2 从第二代数字图书馆的角度出发

笔者在《数字化时代的图书馆服务创新》一文中指出,“支持信息的普遍存取、分布式信息存储与管理以及集成化信息服务”<sup>[2]</sup>是第二代数字图书馆的主要特征。第二代数字图书馆以分布信息处理与集成检索为特征,强调开放式的信息描述与组织,以用户跨类型、跨载体、跨时空地发现和获取信息为着眼点,可以称为“基于资源的数字图书馆”。而复合图书馆的资源组成也包括了数字图书馆的资源。在第二代数字图书馆的技术支撑方面,复合图书馆依然可以大力借鉴。

复合图书馆是诸多高新技术综合应用的产物,因此,从第二代数字图书馆的角度出发,复合图书馆的技术支撑体系可以由下列技术来组成:数字化技术、数据挖掘技术、超大规模数据库技术、网络技术、多媒体信息处理技术、分布式处理技术、安全保密技术、数据可靠性技术、数据仓库与知识发现技术、信息抽取技术、基于内容的检索技术、信息压缩和传送技术、自然语言理解技术。

##### 4.3 从复合图书馆自身特色的角度出发

前文已经论述过,复合图书馆不仅是对传统图书馆与数字图书馆两者特色与精髓的有机糅合,更不乏自身的特色。复合图书馆的一个目标就是要将知识发现、资源定位、查询请求、资源传递集成到一块。以 AGORA 这一复合图书馆的原型项目为例,它就是要解决这样的问题:即希望对各种资源与服务实行有效整合;通过统一的界面集成化地检索不同介质和不同格式的分布式的信息;一次性完成信息发现、定位、提出请求、文献递送的全过程,建立面向用户的、个性化的信息服务机制。

技术的使用要以维护与强化复合图书馆的特色为根本出发点。因此,站在这一角度上分析,复合图书馆的技术支撑体系至少需由以下几部分组成:①推送技术。这是一种按照用户指定的时间间隔或根据发生的事件把用户选定的数据自动推送给用户的计算机的数据发布技术。②搜索技术。在复合图书馆中,由于面对海量的印本资源与电子资源,搜索引擎的重要性可想而知,人们在这方面已取得突破性的进展。通过使用先进知识组织技术,尤其是近阶段来本体(ontology)在知识组织体系上的应用,使得检索者不仅可以实现传统的全文检索,更是逐渐向概念检索等高智能性的检索过渡。③跨库集成检索技术。用户在统一的界面输入检索关键词,就可以同时检索多个数据库,跨时空限制地存取。④信息传输技术。信息传输技术为网络的数据高速传送、存取提供了保证,信息的安全防护及管理技术更是保障复合图书馆正常运行不可缺少的。

#### [参考文献]

- [1] Bruce J Reid, William Foster. Achieving Cultural Change in Networked Libraries [M]. Hampshire: Cower Publishing Limited, 2000. 12-123.

● 师宏睿 (中共宁夏回族自治区委员会党校图书馆, 宁夏 银川 750021; 中国科学院文献情报中心研究生院, 北京 100080)

# 信息世界 1、2、3 及其于图书馆学情报学本体论重构的意义

**【关键词】** 波普尔世界 1、2、3; 信息世界 1、2、3; 文本; 阐释; 认知; 信息三态; 人文本体论; 图书馆学; 情报学

**【摘要】** 在以往研究的基础上, 对两种本体论(知识本体论、信息本体论)意义上的世界 1、2、3 划分做了比较分析, 重点讨论了信息世界 3 对于图书馆学情报学本体论重构的科学能指意义。

**【中图分类号】** G250; G350

**【文献标识码】** A

**【文章编号】** 1005-8214(2005)06-0011-03

世界著名科学哲学家卡尔·波普尔关于三个世界划分的理论是一个具有世界性影响的知识哲学理论。20 世纪 80 年代初, 英国著名图书馆学情报学家布鲁克斯将之引入图书馆学情报学领域,<sup>[1]</sup> 引起了广泛关注和长时期的激烈论争。为助益于业界对这一理论的认识与应用, 笔者曾从信息哲学的视角对之做了一些研讨,<sup>[2]</sup> 本文是对这种研讨的进一步展开。

## 1 两种本体论意义上的世界 1、2、3 建构

### 1.1 波普尔的世界 1、2、3 划分: 基于知识本体论

在波普尔之前, 人们已经习惯于将世界划分为非此即彼的两个部分——物质世界与精神世界或曰客观世界与主观世界。而波普尔却别出心裁地将全部存在和全部经验划

分为三个世界, 即: 世界 1——物质客体和状态的世界; 世界 2——意识形态和各种各样的主观认识世界; 世界 3——包括整个客观知识在内的由人类所创造的文化世界。(注: 波普尔关于三个世界的划分的图示参阅文献 [3])。

在讨论波普尔世界 1、2、3 理论的科学能指意义方面, 笔者认为有几个基本点必须强调:

(1) 波普尔世界 1、2、3 理论是立足于知识(精神)本体论视角的一种世界观, 不能简单地将其之“哲学基本问题”论域的“二元论”或“多元论”混为一谈。在坚持哲学基本问题的物质一元论的前提下, 人们有可能更必要对一切领域的问题进行二元甚至多元化的思考与认识。<sup>[4]</sup> 在这个意义上, 当代知识经济领域有人提出的将世界划分为物质世界、心灵世界、经历世界<sup>[5]</sup> 的主张可以看作是对波普尔世界 1、2、3 理论的继承与发展。

(2) 从信息哲学的角度审视波普尔世界 1、2、3, 可以认为, 它实际上已经从知识的本体论视角对物质的直接存在性(即世界 1)和间接存在性(即世界 2、3)做了明确的划分。信息哲学认为, 知识作为人的主观能动性 & 客观必然性辩证结合的产物与结晶, 属于人类认识(精神)的范畴, 因而在一定的语义上等同于信息体系中的文化信息。<sup>[6]</sup> 这样, 波普尔的世界 2、世界 3 的价值便恰恰在于它们在实质上具有了信息世界的意义。当然, 这种意义更多地应在名物论的范畴(相当于语言学上的名词性概念)上来理解。

版社, 2002. 8-22.

[2] [8] 孙坦. 数字图书馆理论与发展模式研究 [D]. 中国科学院博士学位研究生学位论文, 2000. 9-43.

[3] Pinfield S. Hybrids and clumps [J/OL]. <http://www.ariadne.ac.uk/issue18/main>, 1998-12-10.

[4] HiLiFe. Interoperability [EB/OL]. <http://hylife.umn.ac.uk/toolkit/Interoperability.html>, 2002-01.

[5] Pinfield, etc. Realizing the Hybrid Library [J/OL]. <http://www.dlib.org/dlib/october98/10pinfield.html>, 1998-10.

[6] Judith Pearce. The Challenge of integrated access: the Hybrid library system of the future [EB/OL]. <http://www.nla.gov.au/nla/staffpaper/jpearcel.html>, 2000-02.

[7] Rusbridge C. Towards the hybrid library [J/OL]. <http://mirrored.ukoln.ac.uk/lis-journals/dlib/dlib/dlib/july98/rusbridge/07rusbridge.html>, 1998-07.

[9] 田捷. 数字图书馆技术与应用 [M]. 北京: 科学出

[10] Lorcan Dempsey. AGORA [EB/OL]. <http://www.ukoln.ac.uk/dlis/models/models6/iap/>, 1998-12.

[11] Oppenheim C, Smithson D. What is the hybrid library? [J/OL]. [http://jis.sagepub.com/cgi/searchsortspec=relevance&author1=Oppenheim&fulltext=+hybrid&pubdate\\_year=1999&volume=&firstpage=](http://jis.sagepub.com/cgi/searchsortspec=relevance&author1=Oppenheim&fulltext=+hybrid&pubdate_year=1999&volume=&firstpage=), 1999, (25): 97-112.

[12] 孙坦. 数字化时代的图书馆服务创新 [J]. 中外科技信息, 2002, (4): 18-20.

**【作者简介】** 黄国彬, 中国科学院文献情报中心 2003 级硕士生, 发表论文 2 篇; 孙坦, 中国科学院研究生院硕士生导师, 文献情报中心副主任, 研究馆员, 发表论文 20 多篇。

**【收稿日期】** 2005-01-09 **【责任编辑】** 王 岗