

GUANCANG ZIYUAN JUHE YANJIU YU
SHIJIAN JINZHAN

馆藏资源聚合研究与 实践进展

 《图书情报工作》杂志社 编

名家视点 第6辑

馆藏资源聚合研究与实践进展

《图书情报工作》杂志社 编

海洋出版社

2015年·北京

图书在版编目 (CIP) 数据

馆藏资源聚合研究与实践进展/图书情报工作杂志社编. —北京: 海洋出版社, 2015. 5

ISBN 978 - 7 - 5027 - 9137 - 7

I. ①馆… II. ①图… III. ①图书馆 - 藏书建设 - 文集
IV. ①G253 - 53

中国版本图书馆 CIP 数据核字 (2015) 第 076913 号



责任编辑: 杨海萍

责任印制: 赵麟苏

海洋出版社 出版发行

<http://www.oceanpress.com.cn>

北京市海淀区大慧寺路 8 号 邮编: 100081

北京旺都印务有限公司印刷 新华书店北京发行所经销

2015 年 5 月第 1 版 2015 年 5 月第 1 次印刷

开本: 787 mm × 1092 mm 1/16 印张: 23.75

字数: 407 千字 定价: 48.00 元

发行部: 62132549 邮购部: 68038093 总编室: 62114335

海洋版图书印、装错误可随时退换

《名家视点丛书》编委会

主 任：初景利

委 员：易 飞 杜杏叶 徐 健 王传清

王善军 刘远颖 魏 蕊 胡 芳

袁贺菊 王 瑜 邹中才 贾 茹

刘 超

序

由《图书情报工作》编辑部编选的《名家视点：图书馆学情报学理论与实践丛书》第6辑即将由海洋出版社出版发行与广大读者见面。这是一件值得高兴的事情。从期刊的角度，这是编者从大量的已经发表的文章中精心挑选出来的专题文章，虽然均在本刊发表过，但以专题的形式集中出版，是期刊内容与论文内容的一种增值，体现期刊价值的再利用；对作者而言，这是另外一种传播途径，增强研究成果再次被阅读、被利用的机会，实现论文再次得到关注和充分利用；对读者而言，通过专辑而阅读到多篇同一专题的文章，可以高效率地了解 and 跟踪该领域的研究进展，深化对该领域的认识，对于开展深度的研究或应用到实践工作奠定良好的基础。

本专辑共有4册。第一册是《机构知识库的建设与服务推广》，共收录32篇文章，涉及到机构知识库从基本概念、政策、技术、应用、服务的各个方面，也基本涵盖了机构知识库建设与服务的各个方面的问题，也是有关机构知识库国内重要作者研究成果的大汇聚。机构知识库作为开放获取的重要内容和学术机构自主知识资产的管理与服务系统，是知识管理的重要体现形式，也是图书馆业务与服务新的增长点，具有良好的发展前景和战略意义。对图书馆而言，开发、管理、维护机构知识库并提供基于机构知识库分析的情报分析与科研布局咨询，对图书馆业务与服务的转型发展具有十分重要的意义。

第二册是《移动图书馆服务的现状与未来》共收录37篇文章，涉及移动互联网用户阅读行为、移动图书馆服务模式、移动图书馆服务质量控制、国外移动图书馆服务实践进展、移动图书馆需求与评估等方面。移动图书馆服务在国内图书馆界研究成果不少，学界和业界也高度认同，但由于收到诸多因素的制约，实践上的发展并不够普及和深入。随着移动互联网技术的发展和相关设施的普及，移动图书馆建设仍然是一个值得重视并加大投入的一个领域，其发展前景将十分广阔。

第三册是《馆藏资源聚合研究与实践进展》共收录32篇文章，涉及馆藏聚合模式、数字资源语义关联、关联数据与本体、协同推荐、知识图谱、面向下一代的知识整合检索等。馆藏资源聚合是一个前沿性命题，也是图书馆从资源建设走向基于资源的挖掘与服务的必然过程。这些方面的研究对于深

度地利用馆藏数字资源，实现馆藏资源价值的最大化，具有十分重要的现实意义和应用前景。

第四册是《知识网络研究的进展与创新》共收录 31 篇文章，涉及科研合作的网络分析、共词分析、主题演化分析、学科知识结构探测、研究热点聚类、科研合作网络等，体现了学界业界对这些领域的最新探索和应用性研究成果。为科研提供深度的前沿热点揭示和发现服务，对图书馆服务能力的提升具有重大的意义。图书馆（特别是大学图书馆和专业图书馆）需要加大这一领域的研究、研发和应用的投入，加快图书馆向知识服务的转变。

虽然本专辑的这 4 本书只是从《图书情报工作》近年来发表的文章精选出来的，但也可基本上代表国内学界对相关问题的最新研究成果和图书馆界实践上的探索与创新，具有学术上的引领和实践上的示范作用。尽管研究者还不够多，研究水平也还有待提升，实践应用也处于探索阶段，但也能显示作者们对这些领域的贡献以及潜在的广泛应用价值。

期待这些研究成果能通过这一专辑的出版，对推动国内的学术和实践产生应有的作用，引起更多的图书情报机构的重视，引发更多的研究人员的后续研究，并不断走向深化。在此也感谢所有作者的智慧和贡献，感谢海洋出版社的倾心出版，感谢编辑部同仁所付出的努力。

初景利

《图书情报工作》杂志社社长、主编
中国科学院文献情报中心教授，博士，博士生导师

2015 年 4 月 23 日 于北京中关村

目 次

专 题 篇

- 专题1:馆藏资源语义化与多种方法的融合研究 邱均平(3)
- 分面组配思想在馆藏资源知识组织中的借鉴研究 邱均平 张 聪(5)
- 信息计量学视角下的数字文献资源语义化关联揭示
..... 王菲菲 邱均平 余 凡 赵蓉英(17)
- 基于作者共被引的馆藏资源深度聚合模式与服务探析
——以 CSSCI 中图书情报领域本体研究为例 邱均平 周 毅(31)
- 专题2:数字资源语义互联模式研究与互联方法实例 牟冬梅(43)
- 数字资源语义互联的模式及其比较研究 牟冬梅 张艳侠 黄丽丽(45)
- 基于关联数据的数字资源语义互联模式研究
..... 黄丽丽 牟冬梅 张 然(55)
- 关联数据的语义互联应用研究
——以 VIVO 为实例 张艳侠 齐 飞 毕 强(65)
- 领域本体映射的语义互联方法研究
——以药物本体为例 王丽伟 王 伟 高玉堂 刘宏芳(75)
- 专题3:面向下一代知识服务的图书馆资源整合与检索研究 汪东波(86)
- 略论图书馆资源整合与检索系统的发展
——以国家图书馆“文津”搜索系统为例 ... 申晓娟 李 丹 王秀香(88)
- 基于地方志资源的知识聚合服务系统构建 李春明 萨 蕾 梁蕙玮(98)
- 关联数据在馆藏书目组织中的应用进展研究 赵 悦(106)

理 论 篇

- 国内馆藏资源聚合模式研究综述 赵蓉英 王 嵩 董 克(119)

图书馆、档案馆和博物馆资源整合的发展趋势

- 基于 ICA、IFLA 和 ICOM 历届会议主题的研究 胡心悦(130)
- 国外公共数字文化资源整合管理体制模式及其适用性研究
..... 李金芮 肖希明(144)
- 面向用户的高校图书馆网站服务资源重组模式研究 张 晗 孙 翌(160)
- 数字转型背景下的我国数字档案资源整合与服务研究框架
..... 安小米 白文琳 钟文睿 孙舒扬(171)
- 社会化标注系统资源多维度聚合机理研究 ... 杨 萌 张云中 徐宝祥(186)
- MOOC 教育资源语义化关联研究 ... 陈大庆 丁 培 叶 兰 胡燕蕊(198)
- 馆藏资源元数据的关联网络结构探析:面向 FRBR 解构的视角
..... 成 全 许 爽(208)
- 高校图书馆馆藏资源协同推荐系统研究 邱均平 张 聪(219)

实 践 篇

- 基于多源网络学术信息聚合的知识图谱构建研究 张 洋 谢卓力(233)
- 基于关联数据的非遗数字资源聚合研究 全召娟 许 鑫 钱佳轶(259)
- 本体与关联数据驱动的图书馆信息资源语义整合方法及其测评
..... 欧石燕 胡 珊 张 帅(271)
- 以 UMLS 语义命题为基础的医学信息资源聚合 郭少友 李庆赛(289)
- 基于 FCA 和异构资源融合的本体构建研究
..... 邱 璇 李端明 张智慧(303)
- 基于多特征融合的中文情感分类方法研究 甘小红 张兆年(316)
- 融合语义关联挖掘的文本情感分析算法研究 明均仁(327)
- 高校图书馆学科书评网络资源深度聚合服务探析 李 明(337)
- 基于特征项的文献共现网络在学术信息检索中的应用
..... 丁 洁 王曰芬(345)
- 从大众分类到层次式资源组织体系
——利用聚类信息构建标签树 罗鹏程 陈 翀(358)

专 题 篇

专题 1: 馆藏资源语义化与多种方法的融合研究

序

邱均平

男, 1947 年生, 教授, 博士生导师, 珞珈杰出学者, 武汉大学中国科学评价研究中心主任, 研究方向为科学计量学与信息计量学、科学评价与知识管理, 是我国著名情报学家和评价管理专家、文献计量学的主要奠基人之一。主持或完成国家和省部级课题 37 项, 获国家社会科学基金重点项目优秀成果奖等 60 余项学术奖励; 出版著作 55 部, 在国内外重要期刊发表论文近 500 篇。从 2004 年起, 每年主持研发的“四大评价报告”受到广泛认可, “武大版”评价品牌成为国内外著名的评价品牌之一。E-mail: jpqiu@whu.edu.cn。

馆藏资源语义化是语义网实现过程的重要组成部分, 图书馆也是语义化理论与实践研究的理想阵地。语义网的提出至今已有十几年时间, 人们对语义网环境下的生活充满期待, 众多机构和个人将身边的信息资源发布成语义信息, 使之成为语义网的一部分。信息资源语义化的形式有很多种, 凡是人们掌握的知识通过先进技术转化成机器能够理解的语言, 都可认为信息被语义化了, 所以发布语义信息的途径不仅仅是构成本体或关联数据。

在承担国家社会科学基金重大项目“基于语义的馆藏资源深度聚合与可视化展示研究”的过程中, 我们利用其他学科的方法和理论, 试图挖掘馆藏资源语义化的途径和方法, 本专题的 3 篇文章就是馆藏资源语义化与多种方法的融合研究。

《分面思想在馆藏资源知识组织中的借鉴研究》一文在总结馆藏资源知识组织的现状与不足的基础上, 利用分面组配、知识组织和本体思想的共通性, 将分面组配的方法应用到馆藏资源知识组织的过程中。《信息计量学视角下的数字文献资源语义化关联揭示》一文围绕发文共现与耦合、引文共现与耦合、发文—引文共现 3 个层面对数字文献资源中的元数据概念以及其中的计量语义关系进行了全面揭示和扩散推演, 正是利用了信息计量学的理论和分析方

法，才能挖掘出数字文献资源的深层语义关系。《基于作者共被引的馆藏资源深度聚合模式与服务探析》一文同样利用计量学中的作者共被引分析，探索基于作者共被引的馆藏资源聚合模式，从而找寻馆藏资源深度关联的途径。

利用分面组配的思想，可以更好地规范本体，从而实现语义化；利用信息计量学的相关方法，可以挖掘深层次的语义关系。无论使用哪种方法和途径，需要达到的目标是一致的，就是实现馆藏资源的语义化。我们国家社会科学基金重大项目的语义化方面的后续研究期待与同行在语义化方法和途径层面上的交流，以使理论和应用成果更加丰富。

分面组配思想在馆藏资源知识组织中的借鉴研究^{*}

邱均平 张聪

(武汉大学信息管理学院)

1 引言

馆藏资源是图书馆所有信息资源的集合,包括以实物形态存在和以数字形态存在的信息资源。在网络时代来临之前,馆藏资源主要以实物形式(纸质的书籍、论文集、期刊、报纸等)存在,馆藏资源总数和增长速度都很有限,用户在利用馆藏资源时,只存在找得到或者找不到所需的馆藏资源的情况。而在当今这个被号称“信息爆炸”的时代,不仅仅实物形式的馆藏资源增长迅速,以数字形式存在的、增长更为迅速的馆藏资源在现代图书馆中占据着越来越重要的位置,用户在利用数量众多、形式多样的馆藏资源时不再只是找得到或者找不到相关的信息资源,而往往是在众多的相关信息资源中显得无从下手。传统的信息组织已经很难胜任准确地为用户提供其所需的信息的工作,这时对馆藏资源的组织就需要深入到蕴含在馆藏资源之中的知识单元或馆藏资源之间的知识因子当中,这也是知识组织的主要内容。

知识组织这个概念最早由英国著名的分类法专家 H. E. 布利斯在 1929 年提出^[1],但直到目前关于知识组织也仍然没有一个统一的定义。虽然在概念表述上有所出入,但王知津、英国情报学家布鲁克斯以及马费成基本都认为知识组织是针对知识单元(知识因子)以及知识单元之间的联系进行的^[2],笔者同样持这一观点。知识组织是信息组织的高级形式,是为了减少信息爆炸所带来的信息淹没感,更为快捷地为用户提供其所需的信息或知识。

^{*} 本文系国家社会科学基金重大项目“基于语义的馆藏资源深度聚合与可视化展示研究”(项目编号:11&ZD152)和中央高校基本科研业务费专项资金“基于资源本体的语义信息检索研究”(项目编号:2013104010201)研究成果之一。

2 现状及不足

传统的馆藏资源知识组织主要以分类法和主题法为主（主要是列举式分类法和叙词法），是以文献为单位进行的，也可以说是文献组织。现今随着技术的发展和知识组织重要性的日益突出，研究者提出将各种新技术，例如面向对象技术、简单知识组织系统 SKOS、主题图、本体等引入到知识组织中，但基本上仍然是围绕列举式分类法和叙词法的。

王军指出元数据是当今数字图书馆中最重要的信息资源，但却没有对其进行像针对馆藏资源那样的目录组织和索引，因而“肢解了隐藏在其中的知识系统”^[3]。针对这种情况，他使用 Oracle 9i 中的面向对象技术，集成了分类法、主题法和元数据构造了数字图书馆的知识组织原型系统 VISION。刘丽斌等通过计算机自动转换的方式，用 SKOS 重新描述《中国分类主题词表》^[4]，贾君枝等人则同样通过自动转换的方式将《汉语主题词表》（以下简称《汉表》）转换成 XML 文档^[5]，欧阳宇等人试图通过本体实现《中国图书馆分类法》（以下简称《中图法》）的类目可视化^[6]。

诸如此类的研究很多，大部分都是依靠国外现成的工具（或者是可视化编辑器，或者是 API 工具包），换一种载体重新存储列举式分类法、叙词表和元数据中的语义。列举式分类法是传统书目排架和目录组织环境下的产物，是一种为方便书目排架和查找而设计的线性结构，在数字图书馆环境下已不再适用^[7]。叙词法描述的是网状联系，但其描述的基本对象仍然以文献为主，无法深入到馆藏资源的知识单元中进行描述。

3 分面组配与本体

针对列举式分类法和主题法在揭示馆藏资源知识单元的维度和深度的不足，笔者提出借鉴分面分类法中的分面组配思想对馆藏资源进行知识组织，同时使用 OWL 本体来存储分面分类知识库。分面组配在描述馆藏资源的深度和广度方面的能力大大超过现今广泛使用的列举式分类法和叙词法，而本体则被广泛用于知识组织领域。

3.1 分面组配

分面组配是指运用分析-综合的方法从概念的多个角度出发对其进行组配描述。印度的图书馆学家阮冈纳赞首次明确提出了分面分析的学说，并编制了冒号分类法，将分面分析和分面组配的思想用于图书馆书目组织中^[8]。冒号分类法用于传统图书馆的实践并不成功，这很大程度上是源于冒号分类

法网状的联系不适合用于对实体图书进行线性排架。由于空间限制,传统图书馆中一本图书只能按照一个顺序进行摆放、固定在一个位置,除非在多个不同的位置为同一馆藏资源按照不同的顺序提供多个副本,而这毫无疑问与图书馆有限的资金和空间相矛盾。但这并没有影响分面组配的思想在知识组织领域发挥的空间,甚至早在1953年的Dorking会议上冒号分类法就被创造性地确认为知识组织的基础^[9]。在当前的Web 2.0时代,超链接的存在使得资源访问可以跨越空间的限制,给分面组配思想带来了前所未有的机遇。

淘宝网作为中国最大的C2C电子商务交易平台,在其商品搜索中也引入了分面组配的思想。淘宝搜索的分面思想体现在:用户输入一个关键词并进行检索之后,淘宝搜索分析这个关键词与何种商品相关,并列出这类商品的各个面供用户进行组配筛选。

图1中是笔者输入“手机”之后,淘宝搜索提供的组配面。从图中可以看出,淘宝搜索中与手机相关的面包括“价格区间”、“品牌”、“主屏尺寸”、“CPU核心数”、“操作系统”、“网络类型”6个面。并且,不仅这6个面可以相互组配(面与面之间是“且”的关系),同一个面中也可以进行多选(同一个面之间各个点是“或”的关系),这十分类似于冒号分类法中“轮”的概念。淘宝搜索没有选择对所有的商品的特性做类似冒号分类法中5个范畴的抽象,而是选择根据不同商品的特性选择给以更加具体的面,这源于普通用户对简明性和易用性的要求。

价格区间:	400元以下	400-900元	900-1700元	1700-3400元	3400元以上								
品牌:	小米	三星	诺基亚	苹果	华为	联想	天语	中兴	酷派	索尼	HTC		
	vivo	OPPO	摩托罗拉	魅族	大显	尼凤凰	西维	+多选					
主屏尺寸:	1.8英寸	2.0英寸	2.2英寸	2.4英寸	3.5英寸	4.0英寸	4.3英寸	4.5英寸	4.7英寸				
	5.0英寸	5.5英寸	5.7英寸	+多选									
cpu核心数:	四核	双核	单核	无	四核+四核								
操作系统:	ANDROID	无操作系统	iPhone	Symbian	Windows phone	阿里云操作系统	Windows Mobile						
	黑莓	Linux	OMS	meego	Palm	+多选							
网络类型:	WCDMA(3G)	TD-SCDMA(3G)	GSM	CDMA2000(3G)	双模(GSM/CDMA2000)	GSM/WCDMA/LTE							
	GSM/WCDMA/CDMA/CDMA2000	CDMA	双模(GSM/CDMA)	GSM/TD-SCDMA/WCDMA	小灵通								
	TD-LTE(4G)/TD-SCDMA/GSM	+多选											

图1 淘宝的商品分面^[10]

3.2 本体

根据T. Gruber的定义,本体即对于共享概念体系的明确而又详细的形式化说明^[11]。在OWL本体中,“形式化说明”是通过一系列的包含主语、谓

语、宾语的三元组 (triple) 或者称之为称述 (statement) 来表示的, 而充当主语、谓语、宾语的元素就是“共享概念”。本体中的每一个称述就是知识因子, 而知识关联则通过多个称述之间共同拥有的元素 (主语、谓语、宾语) 体现出来。由此可见, 本体十分适合用于知识组织、绘制知识地图。

当前将本体用于馆藏资源知识组织的研究仍然摆脱不了列举式分类法或叙词法, 本体的类、子类、实例之间网状的联系可以完全兼容分类法或叙词法中的范畴索引树状的联系, 这使得图书馆学的研究者们很容易就联想到利用拥有更强大的语义表达能力、并且更适合数字图书馆环境的本体对传统信息组织工具进行重新包装, 将分类法、叙词法映射到本体中。映射之后的结果通常是分类法或叙词法的范畴索引中的类、子类转换为本体中的类, 叙词法中的用、代、属、分、参、族转换为本体中的属性 (以 OWL 本体为例, 通常转换为 Object Property)。不得不说, 这种映射实际上是十分初级的, 且不谈分类表中的类目是否可以直接转换成本体的类目, 对于列举式分类法, 本体的语义表达能力被列举式分类法的线性的学科体系所限制而不能充分发挥, 本体成为了列举式分类法的另一种展现形式。叙词表中描述的是网状的联系, 但也仅仅只有用、代、属、分、参、族这几种关系, 本体仍然无法发挥其强大的关系描述能力。在 OWL 本体中, 一个类可以拥有多个父类, 一个个体 (individual) 可以从属于多个类 (区别于面向对象编程中的实例), 继承多个类的特征, 这种联系是网状的、组配的联系。

领域本体和上层本体是研究中最常涉及的两种本体。其中上层本体是基于“重用”这样一种思想, 即为了减少重复的工作量而设计的。而当涉及具体的研究领域时, 研究中提到的本体则是领域本体, 即描述当前领域中的概念以及概念之间的关系。本体构建的核心目的之一是为了知识共享, 因此在构建领域本体时应当考虑重用现有的本体尤其是上层本体中的知识, 使得领域本体对上层本体具有兼容性, 减少多个领域本体之间的知识映射的工作量。

3.3 分面组配与本体

分面组配与本体有着天然的联系。分面组配 (比如冒号分类法是通过 PMEST 5 个面) 通过多个维度对一个馆藏进行描述, 不同的馆藏资源通过各个维度彼此联系起来。而在本体中一个个体 (individual) 可以从属于多个类, 可以是多个类的具体化 (specialization)。因此分面组配和本体描述的都是网状的联系, 依据分面组配的思想设计本体不仅可以很好地表达分面组配的关系, 也可以充分发挥本体丰富的语义表达能力。

依据分面组配的思想设计本体时要求将基本面以及其子面分别映射到本

体的基本类以及对应的子类中,在描述馆藏资源时则对馆藏资源进行分面分析,根据分析的结果将馆藏资源设置为其对应的子类的实例,即完成对该馆藏资源的分面组配的描述。

4 分析与应用

由于各个学科、不同的领域都有着自己的独特的知识结构,很难为所有的学科或领域统一设计一个通用的分面体系。基于这样的考虑,阮冈纳赞虽然在冒号分类法中将事物的特性抽象成5个范畴,但他的冒号分类法还是首先按照学科组织,然后在每个学科内部给出具体的面的划分和专有的组配公式。

即使同样都处于网络环境下,馆藏资源尤其是高校图书馆馆藏资源仍然同网络信息资源有着本质的区别。网络信息资源由于其产生自网络,信息质量无法得到保证,并且常常以无序、混乱的形式存在,其用户群体更加广泛、复杂。而馆藏资源主要以书目元数据信息、论文等为主,是经过图书馆员认真筛选、组织之后的高质量、有序的信息资源,并且其用户群体主要以学生、教师、科研人员为主,用户构成相对单一,整体素质相对较高,并且带有明显的学科特性。分面组配用于网络信息组织,其在类目设置时对简明性和易用性的要求较高(例如雅虎直接使用生活用语作为类目,并且一级类目之间有交叉、重叠甚至包含的关系),而在用于馆藏信息资源组织时,则在要求简明性和易用性的同时更加注重严谨性和专业性,以更好地帮助用户从事学术科研活动。

有鉴于此,本文在对类目进行划分时使用教育部最新公布的《普通高等学校本科专业目录(2012年)》中教育学的学科门类作为基本的学科分类体系^[12]。《普通高等学校本科专业目录(2012年)》依次包括哲学、经济学、法学、教育学、文学、历史学、理学、工学、农学、医学、管理学、艺术学12个学科门类,本文只选择其中的教育学学科为例进行研究,并结合教育学学科的特点设计教育学的分面知识体系和领域本体。

4.1 面的划分

对馆藏资源的知识组织不仅要细化到馆藏资源内容特征中的知识因子和知识关联,还要对蕴含在馆藏资源外部特征中的知识关联进行挖掘,即同时用分面组配的思想来组织列举式分类法与元数据中的知识因子和知识关联。本文提出将馆藏资源划分为内容特征面和形态特征面,分别对应着列举式分类法和元数据的内容。

4.1.1 内容特征面的划分 本文将教育学的内容特征面分为教育水平、教育环境、教育内容、教育题材以及空间面,见表1。