

基于语义的馆藏资源深度聚合研究*

贺德方 曾建勋

摘 要 网络环境下,图书馆馆藏资源已逐渐转变为以数字图书馆为主体、紧密面向用户需求的多元化知识资源。本文基于 OPAC、跨库检索、引证关系和元数据的聚合以及相应的整合工具与方法等多个方面对国内外研究现状进行概述,从三个方面进行研究:①基于概念及概念关系、基于引证关系和基于科研本体等对馆藏资源进行语义聚合;②探讨馆藏资源聚合服务所需要的语义描述框架和可视化展示机制;③面向资源主题、知识深化演化、学术社群或科研决策来进行馆藏资源深度聚合。参考文献 23。

关键词 馆藏资源 语义 聚合研究

分类号 G253

Study on In-depth Integration of Library Collections Based on Semantics

He Defang & Zeng Jianxun

ABSTRACT In the networked environment, digital material is becoming dominant in many libraries which are user needs oriented. How to effectively integrate library collection of multi-resources to form structured and orderly resource systems has been a research focus at home and abroad. Based on the summary of status quo of integration of library collections, this paper proposed that library collections should be semantically integrated based on concepts and their relationships, citations and community science ontology. Two important parts of library collection integration, semantic description framework and visualization methods, are explored. And the general idea of how to provide resource content-oriented, knowledge evolution-oriented, research community-oriented and scientific decision making-oriented integration services are also proposed. 23 refs.

KEY WORDS Library collections. Semantics. In-depth integration of resources.

网络环境下,图书馆基于用户需求和文献保障而收集的各类型馆藏文献资源,其内涵和外延已经开始发生变化。从传统的纸质文献为主的采集策略逐渐转向面向用户需求、以数字资源为主体的多元化、开放性知识保障格局,不断加强知识服务功能。一方面,馆藏资源中不同语种、不同载体形式、不同来源的资源之间需要建立有机联系;另一方面,日益丰富的数字资源包含着语义丰富、学科各异、实体多样的知识资源,需要利用多种途径、多种方法和多个界面

来进行揭示、关联和检索,以利于用户对信息资源的有效获取和知识内容的共建共享。因此,如何对馆藏资源进行深层次的揭示、聚合和展示,成为图书情报界关注的热点问题。

1 馆藏资源聚合的国内外研究现状

针对馆藏资源的多元化,学术界一直思考和实践着对多来源的馆藏资源进行整合、集成和融合的问题,以形成结构化、有序化的资源体

* 本文系国家社会科学基金项目“基于引文的知识链接服务研究”(项目编号:08BTQ022)研究成果之一。

通讯作者:曾建勋,Email:zeng@istic.ac.cn

系,从而优化资源利用环境,满足用户需求,实现增值服务^[1]。馆藏资源的聚合研究主要集中在以下方面:

1.1 基于 OPAC 的聚合研究

OPAC(Online Public Access Catalogue,联机公共访问目录)作为图书馆重要的索引机制,是图书馆馆藏资源对外开放和交流的元数据,是图书馆进行资源聚合的基本要素。基于 OPAC 的聚合是通过 Z39.50 协议来实现本馆与外馆的 OPAC 数据库的横向聚合,或者通过在 MARC 记录里增加 856 字段进行揭示或链接书目信息,从而实现不同类型资源的纵向整合^[2]。OPAC 层面聚合研究包括三个方面:①如何将图书馆引进或自建的大量数字资源在馆藏目录中揭示出来^[3];②一个资源分散的单位,如多分校多院系的高校,如何将所有分散资源整合在一个目录体系中^[4-5];③如何通过 Z39.50 协议集成不同图书馆的 OPAC 数据库^[6],等等。基于 OPAC 的聚合研究理论探讨成果丰硕,主要探讨对不同类型馆藏资源进行整合的方法和模式^[7],相关实践也比较成熟,例如,美国国会图书馆的图书馆目录门户提供几百家图书馆的馆藏目录资源检索^[8];MELVYL 联合目录汇集了加州大学 20 多个图书馆的馆藏目录,能够统一界面检索,现在该目录的检索界面已经合并到 WorldCat 中,用户可以选择检索所有书目或者加州大学所有图书馆的目录^[9]。

1.2 基于跨库检索的聚合

基于跨库检索的聚合是多个数据库系统的有机优化整合,是对相关数据库内的数据对象去除重复信息的整合方式。用户不必逐个登录数据库,就可以实现对多个数据库的统一检索^[10]。如方正 Apabi 数字资源平台整合其现有数字资源(含图书、报纸、年鉴、工具书和艺术博物馆),并在此基础上进一步整合、挖掘,从内容层面进行包装,为用户提供统一数字资源发布页面,统一用户管理认证,统一检索入口^[11]。再如 Google 推出的 Google Scholar 有助于用户快速获取网络学术资源,其中涵盖学术出版商、专业

性学(协)会、各高等院校及其他学术机构的图书、期刊论文、报告及其摘要。这些学术资源的排序主要根据其与用户搜寻主题的相关度,其引用情况也会罗列出来。此外,Google Scholar 还会与用户所属机构的图书馆建立链接关系,用户可以直接获取本机构已订购的学术资源,从而真正实现一站式检索^[12]。

1.3 基于引证关系的聚合

基于引证关系的聚合主要是借助于引证关系来揭示文献之间的关联,从而将相关文献聚合成一个相互关联的知识体系。四十多年前,美国科技情报研究所(ISI)创建出版了《科学引文索引(SCI)》,从引文角度揭示期刊文献之间的内在联系。近几年,在迅猛发展的信息技术推动下,ISI 凭借其独特的引文机制和互联网的链接特性,不仅有效地将自身出版的一系列数据库聚合在一起,而且也建立了与其他出版公司的数据库、原始文献、图书馆 OPAC 以及日益增多的网页等信息资源之间的相互链接,构建起一个强大的、基于知识管理的学术信息资源体系^[13]。NEC 研究院于 1997 年创建学术论文数字图书馆 CiteSeer^[14],通过搜索引擎可以获取计算机领域内的学术论文,并提供引文识别符、向量空间等机制来对引文进行描述,如标题、作者、出版年等多种引文的标识。CiteSeer 以此为基础,建立引证文献与被引文献之间的链接关系,并通过图表及可视化方式来显示某一主题文献的时间分布。但是,由于 CiteSeer 是非营利性机构,而且许多文献无法在线获得,所以其学术评价功能还有待改进。

1.4 基于元数据的整合研究

美国博物馆及图书馆服务中心的数字馆藏和内容项目(IMLS Digital Collections & Content, DCC)^[15],在基于元数据的整合研究方面取得很大成绩。该项目主要侧重于资源的互操作问题,基于元数据实现对 IMLS 资助内容资源的一站式检索。现在该项目已不仅仅是整合更多的资源,而是试图在元数据之间建立关系,以便深入揭示资源内容之间的关联,帮助用户实现知

识发现。在基于元数据的整合中,元数据收割协议(Open Archives Initiative Protocol for Metadata Harvesting, OAI-PMH)发挥着重要作用。该框架由开放文档先导(The Open Archives Initiative, OAI)推出,它涉及两类角色:① OAI 数据提供者:建立和维护元数据,并按 OAI-PMH 的要求发布给外界;② OAI 服务提供者:从众多遵循 OAI 协议的机构库中收割元数据,经过加工处理后为用户提供增值服务^[16]。

1.5 整合工具与方法的研究

在对馆藏资源进行整合过程中,人们开发出一些有用的工具,得到广泛应用^[17]。从 1994 年开始,美国斯坦福大学一直致力于图书馆异构数据检索研究,他们所开发出的 InfoBus,将多种不同的数字资源及服务融合在一起,解决 Internet 上异构数据源之间的互操作问题。与此同时,欧洲也开展相关课题研究,如 BIB-LINK^[18], Renardus^[19], CATRIONA^[20], 等等。2000 年,比利时根特大学萨姆堡尔研究小组提出 SFX 技术,这种技术可以把不同来源和不同通信协议的通信源及相关服务融合在一起,从而实现不同类型、不同格式和异地分布数字资源的无缝链接^[21],同时还可以为用户提供上下文相关链接服务。如今,SFX 技术已被嵌入到国外著名数据库中,如 ISI、CSA、OVID 等,在国外很多大学也得到应用^[22]。

总的来说,现有的馆藏资源聚合主要针对文献,基于某种机制或者标准来提供统一的标识和操作,对不同性质、不同来源、不同格式的资源进行描述和链接,使相对独立的资源产生联系,实现对异构资源系统中异质、异类的资源在逻辑上或物理上有机集中,实现资源的全方位整合和一步到位的获取,在一定程度上消除了传统资源与数字资源、异构资源之间人为割裂的零散、孤立的状态,形成一个系统完整的资源集合。不过,现有的馆藏资源聚合着重于从数据和信息的层面进行,导致图书馆服务方式仍以文献查询、原文提供等传统方式为主,对资源内在特征、知识内容的深度揭示,以及基于语义层面的知识资源聚合等方面,都有待于进一

步研究探索。

2 基于语义的馆藏资源聚合方式

在图书情报界,语义主要用于对文献中蕴含知识的多维度揭示,偏重于词汇概念、概念关系和引证关系^[23]。实现馆藏资源深度聚合的基本条件是发现资源之间及其内在的语义关联,语义关系揭示的越丰富,馆藏资源聚合和挖掘的深度与效果就越好。基于语义对馆藏资源进行深度聚合,有助于构建一个内容相互关联、多维度、多层次的资源体系;有助于将不同主题学科、不同内涵外延、不同属性关系的知识内容进行识别、标识和关联,形成集概念主题、学科内容和科研对象实体为一体的立体化知识网络。

2.1 基于概念及概念关系的聚合

在知识组织体系的基础上,通过概念及概念间关系的组织与管理,可揭示馆藏资源之间的语义关联。概念及概念关系的组织与管理包括以下方面:

(1) 基于概念颗粒度。各类知识组织体系通过概念可以实现不同的知识组织目的,但其概念颗粒度有所不同。分类法的概念颗粒度比较粗,适用于学科导航和领域划分等应用;叙词表知识组织的颗粒度相对较小,适用于从概念术语和主题角度组织文献信息;本体的语义关系丰富,概念颗粒度通常比较小,适用于知识挖掘和语义推理等智能检索领域。通过对各类知识组织体系的组合应用,可形成不同颗粒度的语义关联揭示,从而实现不同颗粒度条件下的馆藏资源聚合。

(2) 基于馆藏资源语义分类单元。在传统分类的基础上,采用主题分类与元数据,将馆藏资源在种类、属性、主题等各个方面进行分类描述,储存为不同层次、不同描述属性的语义单元。例如,从主题分类角度,馆藏资源可分为自然科学与社会科学,还可分为数学、物理、化学、天文等;作为知识承载单位的文献,分类单元可以是一本著作、一个章节、一个段落、甚至为一

个语句或术语;资源属性分类可以是数据库、图片库、光盘或网络资源等。在不同分类基础上,研究资源的语义表达方式,可实现基于语义分类单元来揭示资源之间的语义关联。

(3) 基于馆藏资源语义实例。基于本体论,概念是通过具体存在的实例表达的,从实例角度出发,对馆藏资源进行深度聚合,也是一条认识事物、获取知识的重要渠道。在数字馆藏资源中,可通过数据抽取、知识挖掘等技术,获取馆藏资源的实例信息,包括人名、地名、机构名、各类产品型号、药品名称、疾病名称、物种名称,等等。在获取实例后,通过对实例的层次关系进行分类,可揭示实例相关馆藏资源之间的语义关联,通过进一步描述和表达,可从实例角度,将馆藏资源进行聚合。

(4) 基于概念关系。概念关系可分不同的层次,主要包括概念的包含关系,例如分类法的类目等级关系,叙词法的属分关系,本体类的树状结构包含关系等;概念间关系可以根据馆藏资源聚合的需要,研究采用不同的概念关系种类实现馆藏资源的语义表达,例如分类法中的交替类目,叙词表中的相关类目,本体中通用的概念关系与基于领域知识内涵建立的专业概念关系,根据馆藏资源的特征和用户需求,使用不同的概念关系,实现知识的语义表达,揭示馆藏资源之间的关联关系。

(5) 基于概念映射。通过概念映射实现不同知识组织体系的互操作,可以揭示基于不同知识组织体系描述的馆藏资源之间的相互关联,通过映射特征的调控,实现不同图书馆之间的馆藏资源深度聚合。在数学中,映射的基础概念是设 X 、 Y 是两个非空集合,如果存在一个法则 f ,使得对 X 中每个元素 x ,根据法则 f ,在 Y 中有唯一确定的元素 y 与之对应,则称 f 为从 X 到 Y 的映射,记作 $f: X \rightarrow Y$,其中 y 称为元素 x (在映射 f 下) 的像,而元素 x 称为元素 y (在映射 f 下) 的一个原像。在图书馆学、情报学中的映射现象,符合映射的基本数学原理。

2.2 基于引证关系的聚合

文献的引证关系是一个大规模的单向无环

网络,体现文献之间知识的流动和继承关系。利用引证关系可以揭示馆藏资源间的语义联系,对馆藏文献资源进行聚合研究。

(1) 基于引文耦合关系。通过大量的不同时期、不同学科、不同类型的引文耦合计算,可以寻求引文耦合与文献聚合之间的关系,建立文献之间的参考文献集合相似性模型,构建通过引文耦合进行文献聚合的通用方法和模型。

(2) 基于引证关系路径。引证关系网络图中从初始点到终结点中间经过的所有结点的顺序排列集合,在网络图中可能存在多条这样的路径。每条路径具有方向性,可以反映文献引用的一个时间顺序。两篇文献之间的路径数目越多,引文路径总宽度越大,根据引证关系的假设推理,理论上文献之间的引文共性也就越多,文献之间的联系也就越多,相关度也就越大。基于引证关系路径可以对馆藏文献进行语义级的深度聚合。

(3) 基于引证关系强度。对于存在直接引证关系的文献,如何评估它们之间的引证关系强度,需要从引证关系网络中分析与之相关的同引及同被引文献,从多个角度给出综合的评估方法。大多数文献之间不存在直接引证关系,但经常存在各种间接引证关系、同引关系和同被引关系等。通过分析直接和间接引证关系的强度来计算文献之间的相关性,得到一个可量化的引证关系强度,利用聚类和分类理论实现对文献的深度聚合。

(4) 基于扩展引证关系。文献中除了文献主体外,还包含了作者、机构、基金项目等多种科研实体。因此,从描述文献之间的引证关系中,可以形成科研实体如作者、机构、基金等之间的引证关系,揭示出科研实体之间在知识点上的相关性。利用扩展的引证关系可以实现对科研实体的聚合,包括同类的科研实体聚合和异类科研实体聚合。

2.3 基于科研本体的聚合

科研本体是指在传统的规范文档基础上,对各种科研实体,如作者、科研机构、科研活动、科学基金等进行规范化、本体化描述,形成规范

化词典。基于科研本体的语义关联揭示,能够准确客观地描述和表示学术研究领域的科研实体。在此基础上,通过科研实体之间的关联关系,达到对馆藏资源的聚合。

(1)科研实体关联分析。在科研本体中,科研实体之间必然存在多种直接或间接的语义关系,借助于科研本体,可将现实中的科研实体通过各种语义关系关联在一起,从而揭示相应资源之间的语义关联。对于科研实体之间的间接关系,可以采用语义推理等模型技术计算其关联关系,包括从关系路径中推导出现实中实际的语义关系,以及这种语义关系的强度。同时还可以基于关联数据,对馆藏资源进行结构化深度整序,实现文献资源的语义数据发布、智能开放获取,提升资源整体的知识化组织程度。

(2)同类科研实体的语义关联分析。在科研实体关联分析的基础上,对同类实体集合按照它们之间不同的语义关系及强度进行聚合,即具有相同语义关系且强度越大的实体更有可能聚合在一起。首先利用图论知识,研究科研实体之间关系路径的距离和连通性,两个科研实体的关系路径越短并且连通性越好,说明它们之间的相关性越强;其次是科研实体之间语义关系的纯粹性,在语义推理过程中,多种不同的语义关系推理变得复杂甚至还可能导致基于已知信息产生多种不同的推理结果。

(3)多类科研实体的语义关联集成揭示。对不同类科研实体作聚合分析,可以揭示不同类科研实体之间的关联或相关性,比如学者对期刊或基金项目的贡献。与同类科研实体聚合不同,异类的科研实体具有不同的特征和属性,需要利用科研本体中科研实体之间的语义关联关系,在各个共同特征上去计算相关性。通过对科研实体的迭代标注和自动学习,对科研实体间的语义关系进行自动标注,从而解决异类科研实体的聚合问题。对多类实体和实体关系进行标注,并分配合适权重,可以通过计算标注词的语义相关性得到异类实体的相关性。

(4)不同语种之间的语义关联揭示。馆藏资源通常会覆盖多个国家和语言。随着学者、科研机构之间国际交流与合作越来越多,对不

同语言馆藏资源的聚合需求日益突出。利用现有机器翻译技术将文献统一成同种语言再进行聚合的方式虽然具有一定效果,但是还存在翻译不准确、效率低等诸多局限性。可以利用科研本体和双语词典来揭示不同语种资源之间的语义关联,从而实现跨语言的馆藏资源聚合。

3 基于语义的馆藏资源聚合服务实现

3.1 馆藏资源的语义描述框架

从文献深入到知识单元,从单一载体形式到多元表征方式,从单一封闭物理馆藏到开放异构式数字资源分布等,都对馆藏资源组织和利用提出了更高要求。为了实现馆藏资源的深度聚合,首先需要建立语义描述框架对馆藏资源之间的语义关联关系进行描述。

(1)馆藏资源的通用概念模型。考虑到 FRBR(Functional Requirements of Bibliographic Records,书目记录功能需求)是遵循面向对象方法和面向图书馆资源的概念框架,而 FRBR 第二、三组实体则与 FRAD(Functional Requirements for Authority Data,规范数据的功能需求)和 FR SAR(Functional Requirements for Subject Authority Records,主题规范记录的功能需求)相关,可以 FRBR 为核心,在 FRBR\FRAD\FR SAR 三者集成的综合概念框架之上构建馆藏资源的通用概念模型,以实现知识概念与概念之间、概念与实例之间、实例与实例之间、引证关系等多个方面集成。馆藏资源通用概念模型建立在用户与资源之间的语义关联基础上,强调资源组织与用户任务活动的紧密互动,通过用户需求描述与表达等建模研究形成用户知识库,基于科研活动中科研本体构建,围绕馆藏资源数据模型与用户任务之间映射集成等来实现对馆藏资源的语义识别和语义关系的有效描述。

(2)馆藏资源集成描述框架。馆藏资源的集成化概念模型本身是抽象的,具备理论意义,然而要发挥其作用就必须通过设计与之对应的描述机制来落实。由于 FRBR 家族概念模型与 RDA(Resource Description and Access,资源描述

与存取)之间存在固有的描述实现机制,可为馆藏资源对象提供深度描述,因此可以采用 RDA 作为集成化描述框架的基础。主要以 RDA 为核心,特别考虑图书馆集成管理系统中馆藏资源的 MARC 存储现状以及网络资源的描述规范,根据通用模型,基于现实描述层面的元数据互操作(体现为映射转化),通过 RDA 提供的数据元素定义、属性关系描述、注册取值词表的控制等多种描述手段,定制设计不同科研情境下的 RDA 应用纲要。在实现通用描述基础上,增加个性化与跨专业描述集成的基础。

(3)资源语义编码。RDF/SKOS/OWL 语义堆栈是语义网的基础编码表征技术,三者之间可以实现递进式语义过渡。遵循馆藏资源集成描述框架,借助 RDF 对以 RDA 为核心的资源描述框架进行语义特征描述,使用 SKOS 连接和集成规范数据与主题概念体系的表征,并与领域本体 OWL 进行衔接,实现资源语义化的描述。通过关联锁链,实现 RDF 的三元组结构在开放关联数据环境中链接路径聚合,通过利用本体技术对引文进行自动标注,在概念映射基础上实现更高层次的引文网络交融与合并,实现领域知识结构的集成。

3.2 馆藏资源聚合结果的可视化展示

馆藏资源蕴含着海量信息和知识,对馆藏资源进行语义层面的深度聚合后,对其聚合结果进行可视化展示,使信息语义形象化,可以提高信息的可理解和可认知性,从而强化馆藏资源中海量信息的有效利用。海量信息通过可视化手段,把非空间数据转换成视觉形式,具备一定的形象特征,使人的形象思维受到启发,从而揭示出数据的深层次内涵,即以图形、图像展示多维空间,显示出数据之间存在的错综复杂的关系和内涵的潜在信息。

(1)聚合结果的可视化布局。将聚合结果中的各个信息对象按照用户的视觉理解习惯放置在给定的可视区域中。一般来说,语义关系越接近的信息对象视觉距离越近,反之越远。需要根据聚合结果中信息对象之间的语义关系为任意两个信息对象设定合适的视觉距离和表

现形式来进行可视化布局,同时进一步考虑聚合结果在多维特征空间中正确表达为二维或三维空间中的视觉距离。

(2)建立用户深度交互的可视化。为了让用户更有效地获取并理解聚合结果中的数据和内容,需要进一步依据用户需求将这些数据和内容进行细化、泛化、组合、分裂等操作,这些操作需要在可视化图形层面上进行。允许用户对可视化模型本身的参数进行修订,让用户直接参与可视化过程等。因此,需要研究并提供一种能让用户进行深度交互的可视化模型。

(3)多层次展现聚合结果。在对聚合结果进行可视化时,如果能对聚合结果进行不同层次的展现,从粗粒度到细粒度,从总体到细节,能使用户从不同层面对聚合结果进行观察、理解和掌握,并且对这些不同层次的可视化结果进行实时更新,使得用户能及时获得不同层次的可视化展现。因此,需要提供展现不同层次聚合结果的可视化模型。

3.3 基于语义的馆藏资源聚合服务

馆藏资源的聚合服务作为图书馆知识服务的一种有效方式,是指图书馆对馆藏资源进行语义聚合组织,利用对外服务平台,集成各种类型的资源、数据、服务和能力,整合各种信息、内容和应用,并与其他服务有机结合,从而满足用户一体化、知识化和语义化的创新需求。

3.3.1 面向资源主题的聚合服务

首先进行面向馆藏资源聚合的主题识别。在深入剖析知识组织体系(术语表、叙词表、分类表等)语义表示机制基础上,基于引证文献和实体关系对馆藏资源进行无缝检索、关联链接、集成融汇、流程嵌入和量化计算,以规范化的概念体系为主干,以引证关系和实体关系为辅助,从多个维度对资源进行聚合。其次,建立特定领域的知识关系网络。抽取出特定领域的关键特征和属性,如科学术语、科研要素,按照学科概念体系,将抽取出的知识对象组织集成到相关的引证关系网络中,实现对馆藏资源中隐性知识的深层揭示。以知识组织体系中的概念和概念关系为主轴,以引证关系和实体关系为辅

助,来发现和建立馆藏资源的内在聚合关系;利用引证关系,如共词、共作者、共机构等分析,进行语义关系计算与分析,提高概念的量化计算能力,实现知识元相关关系的深度挖掘。在对检索结果进行深入挖掘的基础上,构建检索结果在结构、分布、变化、趋势等方面的总体模式揭示。最后,进行主题关联强度量化计算。通过识别、挖掘分散在文献资源之中的知识对象及其相关关系,对主题关联强度进行量化计算,建立可量化的计算模型,获得更好的统计学解释,可以更为深入地实现对知识对象的关联、集成和可视化表示,促进对非结构化的文献信息进行各种相关的对象关联、计量和分析,实现对海量馆藏文献信息的深层挖掘、辨识、揭示和集成。

在面向资源主题的聚合服务中,将综合知识组织体系、引证关系和实体关系,将馆藏资源中的主题进行实例化,在知识网络系统的基础上,将馆藏资源的聚合过程变成一个基于语义的,能够支持智能检索推理的知识发现过程。

3.3.2 面向知识演化的聚合服务

首先确定知识对象的演化时序路径。知识对象主要包括各类专业术语和科研实体要素。从概念和实体关系的角度,确定以概念为中心、以时序为坐标的知识要素和知识领域演化解析的方法,描绘知识领域中的重要关系结构,发现领域知识演化发展的规律。其次,对学科领域内的各研究主题关系进行演化计算。围绕知识演化追踪问题,从引证关系的知识表示、知识聚类和知识演化追踪出发,分析领域内各种研究主题的演化发展过程,领域内重要科学概念的挖掘方法和发展时序等,确定适用于馆藏文献的概念结构聚合和知识演化分析方法。最后,监测知识领域的变化发展情况,分析知识的结构和演化发展过程,揭示科研过程中的学术研究活动关系,提供信息资源数据关联服务。

在馆藏资源的深度聚合服务中,将主要基于知识组织体系和引证关系构建新型语义关联数据服务。通过概念分析、引证分析实现馆藏资源的动态关联、有机融合、相互嵌入调用,构建基于知识内容的演化聚合模型。

3.3.3 面向学术社群的聚合服务

首先,基于科研实体关系和引证关系,应用学术网络关系挖掘分析的模型及方法,对知识领域内的学术网络关系进行分析,来实现对领域科研实体(包括人员和组织)的结构、活动模式和演化模式的挖掘和揭示。其次,根据引证文献中学术参与者(机构和人员)之间合著关系、引证关系、人员机构变动关系、人员和刊物关系等的发现和挖掘,呈现团队中的学术交流情况,识别和发现领域内社团和社团演化过程,对重要机构、团队、科研人员学术关系网络进行探测优化,及时反映当前科研领域态势,辅助科研决策、合作团队选择、主要机构和人员的确认,更好地为这些科研实体的学术活动提供信息资源保障服务。

在面向学术社群的聚合服务中,主要基于引证关系和实体关系来对学术合作关系的紧密度进行测试,发现、构建并服务于学术社群的发展。

3.3.4 面向科研决策的聚合服务

科研决策包括国家战略决策和科研团队的创新决策两个方面。

在宏观层面,以知识组织体系为基础,从引证关系、科研实体关系等角度把握科研工作的纵向发展脉络,同时从学科领域的概念对比等横向角度,对领域内当前研究热点主题进行判断和识别,发现学科领域的薄弱学科并加以改善,发掘具有较强发展潜力的新兴研究项目,在促进科研工作均衡发展的同时推动在若干个领域取得领先优势,为国家层面的科研管理提供决策支持。

在微观层面,基于馆藏资源的语义聚合方法,为科研创新一线的学科专家或者创新团队提供科研全程服务。通过引证关系和概念关系,对科研工作的热点、前沿和趋势进行识别和计量研究,帮助科研机构确立研究课题、挖掘研究潜力,实现对学科资源的聚合和有效配置。从学术查新、文献获取、知识关联与技术路线图等方面构建集成化科研信息环境,实现学术知识资源的共享、交流与学术竞争力提升,支持科研团队的整体创新需求。

4 结语

实现基于语义的馆藏资源深度聚合是支持并促进知识学习和知识创新的需要。采用先进的知识组织、知识表示、知识聚合和可视化方法,对馆藏资源进行组织和展示,能够形成基于知识关联的新型知识资源集合体,实现跨语言检索、知识导航、知识评价、知识发现、情报计算等更具专业性的深层次服务,更好地满足创新主体对知识信息的深层次需求,从而增进知识理解、促进知识扩散、推进知识创新。

实现基于语义的馆藏资源深度聚合还可以推动信息服务向知识服务转变。通过对馆藏资源的深度聚合,促进其从传统检索查询模式向具有知识导航、自动聚类、自然语言检索和双语查询等功能的智能检索模式转变;从文献信息服务向科研对象实体的关联发现、科技热点监测、科技趋势分析的知识服务转变。

面向馆藏资源的语义描述、识别和语义关系的挖掘和链接,将使不同语种、不同载体和不同来源的各种异构文献信息资源得到基于语义的揭示与组织。知识组织体系建设、引文分析、科研实体关联和可视化展示等相关技术的研究,将极大地提升图书馆知识服务能力,促进馆藏资源的充分利用和效益的充分发挥。

参考文献:

- [1] 崔瑞琴,孟连生. 数字信息资源整合问题研究[J]. 图书情报工作,2007,51(7). (Cui Ruiqin, Meng liansheng. The study of digital information integration[J]. Library and Information Service, 2007,51(7).)
- [2] 李红霞. 图书馆馆藏数字资源整合方案研究[J]. 现代情报,2006,26(8). (Li Hongxia. Study of library resources integration[J]. Journal of Modern Information, 2006,26(8).)
- [3] 王谢宁,杨兴凯. 目录服务在信息资源整合中的应用——目录服务的协议及产品分析比较[J]. 情报杂志,2008,27(7). (Wang Xiening, Yang Xingkai. Application of directory service on information integration[J]. Journal of Information, 2008,27(7).)
- [4] 李育嫦. 高校馆藏数字资源整合现状调查与分析[J]. 图书馆建设,2007(6). (Li Yuchang. A general survey and analysis on digital resource integration at university libraries[J]. Library Development, 2007(6).)
- [5] 庞跃霞,曹丽娟,丁申桃. 高校图书馆馆藏目录整合方法探讨[J]. 图书馆杂志,2006,25(4). (Pang Yuexia, Cao Lijuan, Ding Shentao. Discussion on methods of integrating OPAC of university libraries[J]. Library Journal, 2006,25(4).)
- [6] 章成志,苏新宁. 信息资源整合的建模与实现方法研究[J]. 现代图书情报技术,2005,(10). (Zhang Chengzhi, Su Xinning. Modelling and method of information resources integration[J]. New Technology of Library and Information Service, 2005,(10).)
- [7] 范亚芳,边佳平. 对高校图书馆虚拟馆藏资源整合的分析研究[J]. 图书情报工作,2002,(9). (Fan Yafang, Bian Jiaping. An analysis and study on the integration of virtual library resources at academic libraries[J]. Journal of Library and Information Service, 2002,(9).)
- [8] Library of Congress Online Catalog[OL]. [2011-12-04]. <http://catalog.loc.gov/>.
- [9] MELVYL University of California Libraries[EB/OL]. [2011-12-04]. <http://melvyl.worldcat.org/>.
- [10] Chang C, Lu W. Agricultural cross languages information retrieval schema based on multi-thesaurus mapping. 2008' CCTA[C]. Computer and Computing Technologies in Agriculture II, Volume 1. New York: Springer, 2009.
- [11] Apabi 数字资源平台[OL]. [2011-12-04]. <http://ref.lib.apabi.com/Default2.asp?lang=gb>. Apabi digital resource platform[OL]. [2011-12-04]. <http://scholar.google.com.hk/>.
- [12] Google 学术搜索[OL]. [2011-12-04]. <http://scholar.google.com.hk/>.
- [13] ISI web of knowledge. [EB/OL]. [2011-12-04]. <http://wokinfo.com/>.
- [14] CiteSeer; Scientific Literature Digital Library and Search Engine[EB/OL]. [2011-09-01]. <http://citeseer.ist.psu.edu/>.

tp://citeseerx.ist.psu.edu/.

- [15] Collections in the age of e-research; realizing potential through curation and aggregation [EB/OL]. [2011-08-29]. <http://www.clir.org/dlf/forums/fall2010/22PalmerDLF.pdf>.
- [16] 郭少友. 基于 OAI-PMH 的信息资源整合[J]. 大学图书馆学报, 2005, 23(3). (Guo Shaoyou. Information resources integration based on OAI-PMH[J]. Journal of Academic Libraries, 2005, 23(3).)
- [17] 王浩. 图书馆网络信息资源整合工具——MAP 及其启示[J]. 现代情报, 2010, 30(5). (Wang Hao. Tool of network information resource integration in library-MAP and its enlightenment[J]. Modern Information, 2010, 30(5).)
- [18] OCLC-Scorpion[EB/OL]. [2011-08-29]. <http://www.oclc.org/research/activities/past/or-projects/scorpion/default.htm>.
- [19] Renardus: Academic subject gateway service europe[EB/OL]. [2011-08-29]. <http://www.ukoln.ac.uk/metadata/renardus/>.
- [20] CATRIONA[EB/OL]. [2011-08-29]. <http://www.bubl.bath.ac.uk/BUBL/catriona.html>.
- [21] 王军, 张丽. 网络知识组织系统的研究现状和

- 发展趋势[J]. 中国图书馆学报, 2008, 34(1). (Wang Jun, Zhang Li. Network knowledge organization system: Present status and development trends[J]. Journal of Library Science in China, 2008, 34(1).)
- [22] 蒲筱哥. 图书馆数字信息资源整合技术研究进展[J]. 现代情报, 2008, 28(7). (Pu Xiaoge. The analysis of the technologies of digital information resource integration in libraries[J]. Modern Information, 2008, 28(7).)
- [23] Chang C, Zeng J X, Wu W N. Acquisition of Class Relations for ontology based on language expression and knowledge structure[C]. Computer and Computing Technologies in Agriculture II, 2010: 429-432.

贺德方 中国科学技术信息研究所所长, 博士, 研究员, 博士生导师。通讯地址: 北京市复兴路 15 号中国科学技术信息研究所。邮编: 100038。
曾建勋 中国科学技术信息研究所信息资源中心主任, 博士, 研究馆员。通讯地址同上。

(收稿日期: 2012-01-09)