

数字资源聚合方法融合趋势研究

马鸿佳 李 洁 (吉林大学管理学院 长春 130022)

沈 涌 (吉林大学公共卫生学院 长春 130021)

摘要 本文在对主题词表、本体、关联数据、文献计量、分众分类及社会网络分析法进行特征及优势分析的基础上,从方法融合视角对数字资源聚合方法的融合趋向进行了归总,并进一步梳理了方法融合的具体作用机理及应用领域,以拓展数字资源聚合方法融合研究的新方向。

关键词 数字资源聚合 方法融合 机理与应用分析

Fusion Trend of Digital Resource Aggregation Method

Ma Hongjia Li Jie (School of Management, Jilin University, Changchun, 130022)

Shen Yong (School of Public Health, Jilin University, Changchun, 130021)

Abstract On the basis of the analysis of the features and advantages and disadvantages of saurus, ontology, linked data, bibliometrics, folksonomy and Social Network Analysis(SNA), this paper from the perspective of fusion summarizes the trends of methods amalgamation of digital resource aggregation and further anatomizes the specific mechanism and application fields of them to expends the new directions of methods of digital resources aggregation.

Keywords digital resource aggregation, methods amalgamation, mechanism and application analysis

1 引言

置身于浩如烟海的数字世界中,资源的序化、群聚及语义丰富度等愈来愈影响着人们对海量数字资源的使用,由此,数字资源聚合作为一种新兴的知识组织方法逐渐进入人们的视野并被研究者所重视。“数字资源聚合”是学科知识不断积累、奠基形成的知识理论,从化学领域的分子聚合研究到互联网海量信息聚集的网络用语研究再到图情领域的数字资源聚合研究,历经了漫长的发展时期。其含义也从 Etzionid^[1]提出的有关信息资源聚合的早期定义到姜灵敏^[2]、邓胜利^[3]、杜晖^[4]等学者就海量信息资源聚集与异构信息资源融合两个层面对信息资源聚合的具体阐释,再逐渐过渡到毕强等^[5]就知识关联维度所强调的数字资源聚合从组织语义化向内容语义化的转变发展,邱均平等所提出的^[6]数字资源聚集与融合的简明界定以及贺德方等^[7]所给出的从科研实体到科研内容,再到主题概念的“立体化知识网络”的生动说明。

随着理论研究的扩展,数字资源聚合在方法开发、模型构建及应用研究等方面也不断深入,并逐渐生成新的领域方向。特别是主题词表、本体、关联数据、文献计量、分众分类及社会网络分析等数字资源聚合方法的应用,加速了数字资源向知识资源转化的进程。但单一的知识组织方法难以实现资源在广度、深度、维度方面的深层次开发,由此催生了数字资源聚合方法的融合研究。本文在借鉴已有研究成果的基础上,对数字资源聚合方法融合的发展趋向、作用机理及应用领域进行进一步的归总分析,以拓展数字资源聚合方法融合研究的新方向,从而更好地实现数字资源的再组织与再利用。

2 数字资源聚合方法的融合趋向

数字资源聚合的实现离不开技术与方法的支持。有关数字资源聚合方法的研究,图情领域至今还未有一个较为系统、完整的方法体系,但学者们大都从语义与关联两大主线对数字资源聚合方法进行了

本文系国家自然科学基金项目“语义网络环境下数字图书馆资源多维度聚合与可视化展示研究”(编号:71273111)、吉林大学科学前沿与交叉学科创新项目(编号:2011QY109)的研究成果。

梳理。其中,张云中⁸从基于关联数据、本体等方法的数字资源语义研究和基于文献计量实现关联发现的数字资源引证研究的角度将数字资源聚合方法划分为两大体系;周珊珊⁹在张云中研究的基础上,从资源关联与语义角度指出,数字资源聚合方法形成了包括运用关联数据、本体及形式概念分析等强化语义的知识组织方法和运用网络分析、共现耦合增强关联关系的计量分析方法两大思路;毕强等⁹则根据数字资源聚合由概念聚类、概念关联到知识关联的递进关系提出,数字资源聚合方法既包括通过主题词表、社会网络分析、计量学和分众分类法等实现数字资源组织语义化的方法,又包括通过领域本体、关联数据等实现数字资源内容语义化的方法,这两类方法推动着数字资源语义化从组织语义向内容语义的演进与融合。围绕着这两大主线,主题词表、本体、关联数据、文献计量、分众分类及社会网络分析等数字资源聚合方法形成了互补融合的理论及应用模式,出现了本体与分众分类法的融合研究¹⁰⁻¹⁵,主题词表与本体的融合演进研究¹⁶⁻¹⁸,主题词表与分众分类的融合应用研究¹⁹⁻²⁰,社会网络分析与文献计量的融合应用研究²¹,社会网络分析、本体及关联数据的数字关联应用研究²²,本体与文献计量的馆藏资源语义化应用研究²³,分众分类与社会网络分析的聚合应用研究²⁴以及基于本体、分众分类、社会网络分析方法融合的社会化标注系统研究²⁵。

综合以上研究内容,可发现数字资源聚合方法间的融合研究主要有以下特点:一是以本体聚合方法为中心,分别展开了本体与分众分类、本体与主题词表、本体与社会网络分析、本体与文献计量以及本体与关联数据的理论与应用研究。其中,本体与分众分类的研究最为深入,以张云中与陈开慧为代表的学者对两类方法的聚合机理用象限图给予了直观、形象的阐释。二是出现了多种方法的融合应用研究,例如基于社会网络分析、本体及关联数据方法融合的数字文献资源关联关系应用研究和基于本体、分众分类、社会网络分析方法融合的社会化标注系统中的资源多维度聚合应用研究等。三是方法融合间的作用与应用机理多以方法间的互补为基础,通过方法间的优劣协调或是借助能够抵消其方法软肋的相应工具和技术来进行深层次的方法融合应用研究。四是方法融合研究的层面不断扩展,随着方法融合从理论到实践有关研究成果的不断丰富,融合思想越来越受到重视,其正影响着数字资源聚合在多层面的实现。

3 数字资源聚合方法融合机理与应用

大数据时代的到来使数字资源聚合也面临着巨大的挑战。如何实现海量、分布、异构数字资源的有效聚合已成为数字资源聚合方法研究中的重中之重。主题词表、文献计量、本体、关联数据、社会网络分析及分众分类法在数字资源聚合中各有特色,在推动数字资源聚合由理论到实践的转变过程中各自发挥着不同的作用。但单一的方法难以解决大数据环境下所面

临的有效聚合问题,这就需要数字资源聚合的不同方法能够相互结合,形成取长补短式的融合互补合力,推动数字资源聚合更好地朝向深度聚合与广度关联方向发展。实现数字资源聚合方法融合的首要前提是对每一方法特点、优劣势及融合应用的机理进行分析。数字资源聚合方法的优劣如表1所示。

表1 数字资源聚合方法多维度比较分析

如表1所示,多种聚合方法各有所长。其中,主题词表作为兼具结构化与规范化的语义术语关系集合,主要由主表、附表和辅表组成。通过运用具有语义相关性与族群相关性的词或词组以等同、相关、等级等词间关系,对数字资源进行语义化、逻辑化、规范化的描述。主题词表在知识表达上呈线性、一维结构²⁶,是数字资源聚合的一种有效方法,具有结构化、规范化、语义化以及可复用等优点,但也存在着表达模糊、自由度低、词间关系粗糙、扩展性与时效性较差以及非形式化等缺点。文献计量在数学、统计学等理论的基础上通过定量分析文献单元的特征和关系实现对学术发展脉络的评价与预测,借助资源挖掘工具与软件等进行可视化描述,并通过量化结果输出和语义化路径实现文献的多维聚合。作为数字资源聚合研究中常用的统计方法,文献计量在数字资源聚合研究中具有海量信息分析、结果输出量化、多维聚合及可视化呈现等优点,但存在分析内容有限、资源语义挖掘不深以及需要其他工具、技术或方法辅助等缺点。本体(Ontology)是一种能够对数字信息资源进行知识与概念语义揭示,并能对概念与概念关联进行深层挖掘,以实现有效知识表达与概念模型建构的数字资源聚合方法。本体通过类、关系、实例及公理等要素对资源关联关系进行形式化描述,并以网状结构呈多维分布,所形成的概念集合与知识库可自行更新²⁷。总体上,在数字资源聚合中,本体具有形式化、语义关联、概念丰富、开放集成、多维知识化构建的优点,但在使用上也存在着时效性差、易用性与灵活性较低等缺点。

关联数据能够借助 URL、RDF、HTTP 协议等工具实现数字资源在不同数据库、不同研究领域的共享与利用。作为一种无缝链接的数字资源组织方式,其最大优势是能够使分布、异构的数据资源进行无缝连接、整合与模块化集成。关联数据主要通过主体、客体与谓词对数据信息进行描述,并借助资源标识及通讯协议对数据资源进行内外部链接,采用统一标准格式对数字资源及其关系进行管理。总体上,关联数据在数字资源聚合中具有关联有效、语义丰富、易于识别与使用、形式化高以及兼具完整性与准确性的优点,但其无法进行隐性资源关系揭示及深层语义挖掘,且在封闭系统中难以实现数字资源的有效聚合。社会网络分析法^[24]由知识群落、知识关联及知识单元组成,通过社会网络来展现信息资源的关系及属性,随着多维语义聚合的发展,更加注重关系思维的运用。社会网络分析法主要通过网络关联性、中心性及群聚性对知识节点进行定位与划分,通过网络群聚性进行群族分类,通过资源网络链接关系与结构进行关系分析,通过作者、引文、机构等关系图谱进行中心度、凝聚子群、中心势及结构洞分析,通过多维语义揭示作者、流派、引文等特征,并以知识图谱形式呈现文献聚合的从知识单元到知识群聚的多维的、立体的多粒度网络关系^[3]。社会网络分析法具有结构性强、群聚与连通性高、多维立体呈现、语义关联分析及可视化生动呈现等优点,但在用户交流上,SNA 以后台支持为主,缺乏互动性且处理过程复杂,自由度和全面性都较差。分众分类法以用户标注为依据,基于“大众一致意见”^[1,28],根据用户标签点击率进行数字资源分析统计,是一种开放环境下的用户意识体现。分众分类法能够全面、时效地表达用户需求,挖掘潜在语义关联,以平面化、非等级形式进行资源分类,兼具动态性与自由性,又凸显与用户的交互性。作为一种大众化的数字资源组织体系,分众分类法是一种以用户为中心的理念诠释,在聚合过程中具有简单实用、便捷时效、动态交互、分布协同、协作关联等优点,但也存在着规范度与关联度低、结构性与受控性差、语义模糊以及维度单一等缺点。

(1) 数字资源聚合方法融合机理

数字资源聚合的目的在于实现知识发现及知识融合。数字资源聚合方法的融合机理是在吸收已有研究理论及成果的基础上充分挖掘各方法的特点及优劣势,通过方法间的优势重组、互补匹配及弱势改造来强化数字资源聚合方法在数字资源深度聚合与广度关联层面的纵深挖掘,进而促进数字资源向知识资源的深化转变,最终实现知识发现、知识重组及知识利用。数字资源聚合方法融合机理的具体实现层次如下。

① 概念强化层面的方法融合

概念是对客观事物本质性、一般性特征的反映,在知识表达和知识组织的实现中,概念颗粒度越细,越有利于语义表达,概念关系层级越深,越有利于语义推理,最终,越有利于数字资源聚合的实现。概念强化层面的数字资源聚合实现如图 1 所示。

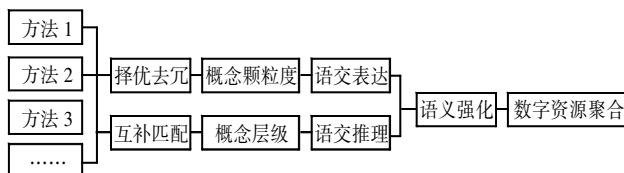


图1 概念强化层面的数字资源聚合方法融合实现图

本体、主题词表、分众分类、文献计量及社会网络分析等聚合方法在概念颗粒度及概念层级与聚类的挖掘上各有优劣,可以通过方法间的择优去冗和互补匹配实现其在概念横向与概念纵向层面的深入揭示,推进数字资源的语义表达与语义推理,最终提升数字资源聚合的纵深发展。一方面,数字资源聚合方法可以通过方法间概念颗粒度的不同情况,通过择优重组更好地发挥方法融合对数字资源聚合的促进作用。随着概念颗粒度由粗到细的变化,所涉及的文献数量将会减少,所要分析的文献单元的关键词数量将会增加,文献单元的概念横向揭示层次则会提高,语义表达程度与知识表示层次的效果也会提升,相应地,也越有利于数字资源聚合的实现。概念颗粒度与知识表示的作用如图 2 所示。

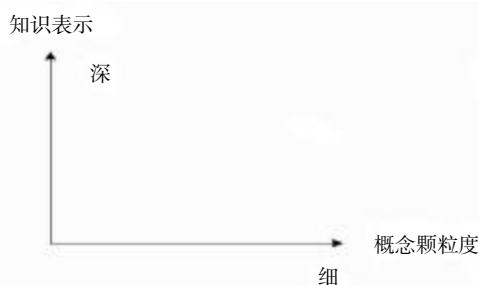


图2 概念颗粒度与知识表示作用机理图

如图 2 所示,受概念颗粒度的粗细影响,知识表示也会由深到浅变化,在知识表达上,主题词表与本体的概念颗粒度都相对较细,语义的表达度也相较丰富,但前者多被应用于从概念主题方面挖掘文献信息,后者则多被用于从知识与语义挖掘层面提升知识组织。所以,利用主题词表与本体的概念横向强化优势,可以发挥两者在数字资源聚合中主题揭示与语义挖掘的双重优势。针对这一理念,一是可以对主题词表进行自动编码,使其可以用于本体的自动检测,与此同时,创建本体制定者所需词表更变的格式标准,通过提升主题词表的语义规范性,提高其语义描述与知识表示能力;二是可以以主题词表为基础对本体进行重新创建,开发一个语义内容更为丰富、概念层级更为多样、语义表达与推理更为精细的本体范例。

另一方面,随着概念层级的细化,概念的纵向揭示会相应提高,语义推理与知识表示程度也会相应深入,相应地也越有利于数字资源聚合的实现。所以,基于不同概念层面对数字资源聚合的不同作用,各聚合方法可以就不同概念层级、概念聚类及关联程度的优劣互补匹配,促进较高概念层级下的数字资源聚合的实现。此种情况下已有的方法融合目前主要有两种:

一是主题词表与本体择优去冗下的概念层级强化实现的方法融合,二是通过本体与分众分类法互补匹配以及文献计量与社会网络分析的互补匹配实现方法融合在概念纵向强化层面的构建。

主题词表和本体都是用以描述特定学科知识领域的知识组织工具,都通过词间关系或等级关系对数字信息进行概念或类的组织与描述,但不同的是,主题词表主要通过“用、代、分、属、参”对概念关系层级进行揭示,其表达较为宽泛与模糊且欠缺逻辑性,本体对概念关系层级的揭示比较多样与规范且更具形式化,其不仅注重语义逻辑而且允许用户自定义概念关系。所以,两者可以通过择优去冗实现概念层面的纵向融合来进一步提升数字资源的概念与语义关系,以促进数字资源的深度聚合。首先,通过主题词表就数字资源从整体层面的概念把握以及本体就概念对象的进一步丰富拓展对概念层级进行更为详细的划分与揭示;其次,通过本体扩充主题词表的概念类目及概念间关系并提升概念及概念关系描述的表达力与推理力;第三,注重有关概念层级划分及概念关联的相关信息,重视方法融合过程中本体的再利用与共享性价值。本体与分众分类是数字资源聚合方法中互补性较强的,本体形式化与语义关联度较高、概念表达丰富、开放集成、结构多维且能够进行知识化构建,其对概念关系的揭示较为充分,而分众分类法规范性与关联度低、结构性与受控性差、语义模糊、维度单一,总体上对概念层级的挖掘相对较差。所以两者可以通过互补匹配进行融合,促进概念层面的纵向深化。文献计量作为定量分析文献单元特征及关系的方法在概念及语义分析上并不深,而分众分类法结构性强、群聚与连通性高,可以进行概念聚类及语义关联分析,并能可视化呈现,所以文献计量与社会网络分析法可以通过互补匹配实现在概念聚类强化层面的概念表达与资源聚合。此外,借助形式概念、概念等工具也可以实现数字资源概念的纵深强化,如形式概念与本体的结合,通过两者优势的结合可以对概念知识进行有效处理以实现知识发现、知识表示及知识获取;又如通过本体与概念格的互补融合,能够借助概念格的层级优势与多重继承关系使本体结构更呈网状,更有利于概念的纵深强化。

②语义强化层面的方法融合

在数字资源聚合的实现中,语义越丰富、表达越规范、语义关联越紧密,越有利于数字资源对象在概念及关系层面的语义揭示、知识表示及知识组织的实现。基于语义强化层面的数字资源聚合方法融合,主要通过两种途径进行:一是基于概念及概念关系的强化提升语义表达与语义推理的目的,运用择优重组或互补匹配实现数字资源聚合就语义强化条件下的方法融合;二是直接基于语义强化层面,通过方法间的互补匹配促进语义表示与语义推理的提高,进而实现数字资源聚合方法的有效融合。其有关语义强化层面的方法融合实现如图3所示。

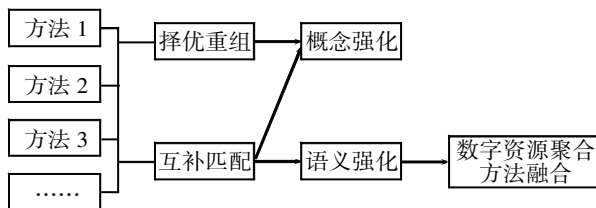


图3 语义强化层面的方法融合实现图

由于途径一在前文已做过介绍,所以本文重点介绍途径二下本体、主题词表与分众分类法通过互补匹配在语义强化层面的方法融合。本体通过“has-a”(整部关系)、“is-a”(属种关系)、“association”(对等关系)及“instance-of”(实例关系)等基础或拓展式的关系采用规范词汇进行语义描述与语义推理,语义表达较为丰富,语义描述较为规范、准确且形式化程度较高;分众分类主要运用用户自定义的自由词通过标签共现进行语义表达,语义描述较为稀疏、模糊,欠缺规范性与严谨性且形式化程度较低;主题词表运用等同、相关、等级等词间关系通过语义相关与族群相关的词或词组,对数字资源进行结构化、语义化和规范化表达,其相较本体虽然在语义揭示与语义推理上有所差距,但较之分众分类,则具有语义描述规范、逻辑与准确等优势。所以,本体与主题词表都可以与分众分类进行融合,利用其语义表达与语义推理优势弥补分众分类在语义方面的欠缺,推动数字资源聚合在深度层面的拓展。在本体与分众分类法的融合上,两者可以在标签成对识别及对比分析的基础上探索并开发标签间的语义关系,通过语义强化共同实现数字资源的聚合;在主题词表与分众分类的融合上,两者在标签关系挖掘的基础上,可以借助相应的语义开发工具,进一步实现语义强化层面的方法融合。

③关联强化层面的方法融合

数据关联是数字资源实现有效聚合与知识组织的重要条件,通过深入分析数字资源的关联关系及语义关系,可以开发数字资源潜在的概念、关联与语义关系,挖掘数字资源的隐性价值,促进知识表示和知识组织的实现。我国目前对数字资源关联关系的揭示,主要通过关联关系发现和关联关系构建来进行,其中关系系数的多少影响着关联的最终效果,关系系数越多,关联的揭示会越广,关联效果就会越好,其作用如图4所示。

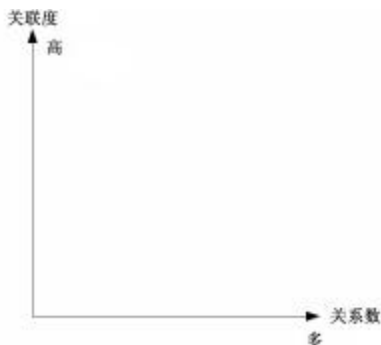


图4 关联强化实现作用图

数字资源聚合的几大方法几乎都具有语义关联的功能,其中,关联数据是最为有效的关联强化方法,其本身不具语义性,无法进行深层语义挖掘与隐性关系揭示,但可以通过建立关联实现语义表达,形式化程度高且易于识别,但其无法进行潜在语义关联挖掘;本体在各方法中的关联度也是较高的,其与关联数据相比,欠缺无缝链接性;其他的方法包括主题词表、分众分类及文献计量等虽具有语义关联作用,但关联度都相对较低。所以,数字资源聚合方法可以通过择优去冗、互补匹配及弱势改造实现关联强化层面的方法融合,促进数字资源的进一步开发,现有的关联及关系强化主要有:一是通过择优去冗,将数据关联与本体构建相结合,一方面促进本体构建中无缝链接关联的实现,另一方面提升关联数据应用中潜在语义关联的挖掘,发挥强强联合的最佳优势;二是通过关联数据与文献计量方法的优势联合,可以更好地对文献单元的内外特征进行归总,深化数据关联;三是通过社会网络分析与本体的结合,发挥社会网络分析的群聚、关联和中心分析作用与本体的语义关联作用,更好地实现关联强化;四是通过方法间关联强弱的互补匹配提高数字资源关联度,如本体与主题词表、分众分类及文献计量的融合,可以将本体良好的关联度挖掘作用于主题词表、分众分类及文献计量的关系揭示实现上,通过拓展关系数及关系层级实现数字资源聚合更好的关联;此外还可以借助耦合关系、概念分析等工具,利用其关联关系优势,对弱势方法进行改造。

④应用强化层面的方法融合

在数字资源聚合的多个方法中,分众分类法的应用性较强,简单易用、动态交互、协作关联、时效便捷而且能够体现用户意志与需求;关联数据语义丰富、扩展性强、易于识别和使用,兼具完整性与准确性;本体与主题词表时效性与扩展性较差、不太易用、灵活性与自由度低;文献计量作为定量分析的数字资源聚合方法,其分析内容有限且语义性与自由度相对较差;社会网络分析法的互动性较差,处理过程复杂,自由度和全面性都较低。针对这几大聚合方法在易用度、实用性、动态性、自由度、时效性、便捷性、扩展性及互动性等的差别,可以看出分众分类与关联数据在应用层面更具优势。

上述方法一方面可以通过方法间的互补匹配,通过提升应用的易用度、自由度及灵活性等实现应用强化。如本体与分众分类方法的结合,利用分众分类法的简单易用性等功能可以提高本体应用的张弛度。本体形式化程度高,对用户的操作使用要求相对较高,在一定程度上不利于其应用的拓展,往往局限于特定领域;分众分类较为简单实用,对用户的专业知识要求不高,适于大范围使用且维护比较简单。本体与分众分类法相结合,可以为本体构建与应用提供新思路,分众分类对用户思想与意志的重视可以为本体构建中由于对用户关注不够而造成的实践与理论相分

离的问题提供解决依据,且其协作互动的特征也有助于发挥本体开发应用中的群策力。对于两者融合的具体实现,一是可以通过概念、属性及实例映射对本体构建进行拓展,二是可以借助分众分类的社会化标注功能为本体构建、维护及应用创造良好的“群众基础”环境。另一方面,也可以借助其他方法,通过劣势改造与优势重组实现应用层面的强化。如文献计量与内容分析法的互补融合,内容分析可以弥补文献计量在量化描述中客观性与环境因素考虑的缺失;再如,分众分类与专家分类法的强强联合,将分众分类自底向上的分类与专家分类法自顶向下的分类结合,既保证应用的灵活性、自由度与易用性,又能保证应用的规范性与准确性。

⑤多维强化层面的方法融合

数字资源聚合方法在多维强化层面的融合主要有两个方面,一是知识组织结构方面的方法融合,即就方法本身知识组织结构的优劣而进行的融合;二是功能实现方面的方法融合,就方法在数字资源聚合过程中不同的功能特点而进行的融合。

首先,在知识结构方面,除主题词表与分众分类分别呈线性与平面性之外,其他方法都成多维性,就这一层面而言,本体与主题词表的融合、本体与分众分类的融合以及分众分类与社会网络的融合都能够通过互补融合,实现数字资源聚合从单一维度向多维立体的转变,促进知识发现与知识组织的实现。其次,在功能整合层面,本体、关联数据、主题词表、分众分类、文献计量与社会网络分析在概念及关系方面、语义与关联方面、内容与形式方面以及应用与组织结构方面各有优劣,通过方法间或是借助其他方法工具的择优去冗、优势组合、互补匹配与劣势改造可以最大化地拓展数字资源聚合的广度与深度。如,本体与分众分类在网络知识开发应用中的多维强化,利用本体的语义关联优势与分众分类的使用、构建优势,对现有网络知识系统进行改进,使其不仅具有社会网络的知识获取与知识组织的特征,还具有语义网的语义表示与语义推理的特征。

(2) 数字资源聚合方法融合的应用发展

自数字资源聚合方法融合的思想提出以来,国内外对方法融合的研究探索在深度与广度上不断得到拓展。其中,方法融合的研究层次从理论到实践不断演进,方法融合的应用范围从数字资源到馆藏资源以及其他具体资源不断深化,方法融合的维度从两两结合到多维结合不断立体。

在概念强化层面,主题词表和本体通过有关概念关系描述方面的集成研究,促进了面向具体应用的概念术语与概念规范的研究并推动了新的元数据模式的开发,此外,两者的融合还被应用于数字图书馆智能导航的概念模型建构;形式概念与领域本体的双向融合,形成了形式概念基础上的领域本体建模理论;本体与概念格的互补融合推动了多本体协同知识地图的产生。基于语义强化层面,出现了融合形式概念

分析强大形式化下的数据分析与规则提取功能、本体的丰富语义关联功能以及集分众分类语义发现功能于一体的语义关系发现模型;融合分众分类“浮出语义”功能挖掘用户网络标签语义关联的本体构建应用;融合本体与关联数据的语义关联功能的异构文献资源整合。基于关联强化层面,本体与关联数据融合下语义关联的充分开发,推动了信息资源语义融合框架的形成,促进了馆藏资源内部间的相互联系以及内外部资源无缝链接的实现,提高了图书馆数字资源的有效聚合与利用,为匈牙利与法国国家图书馆目录与主题标目的产生以及德国国家图书馆档案法的推广应用奠定了基础;关联数据与文献计量的融合促进了知识关联网络的推广及应用。基于应用强化层面,形式概念分析与分众分类的融合推动了有关用户行为与用户偏好的研究;文献计量法和内容分析法的融合促进了定性定量相结合的应用软件系统的开发,其通过对文本信息的内外部特征规律的分析能够为信息服务工作方向提供指引;分众分类与本体的融合实现了知识库构建的可操作性与易用性。基于多维强化层面,本体、分众分类法与社会网络的融合推动了有关标签与标注活动的理论与实例研究;本体、关联数据及文献计量等方法的融合实现了数字文献资源关联关系的多维揭示。

随着数字资源聚合的深入发展,各方法间的融合层面不断延伸,融合机理等理论分析不断深化,融合应用方向也渐趋多样。在知识发现、知识表示、知识组织与知识利用的推动作用下,数字资源聚合方法间的融合研究将继续向数字资源的深度聚合与广度关联方向迈进,各方法间的多维融合趋势也会愈加凸显。

4 结语

随着数字资源向知识资源乃至智慧资源的发展要求转变,单一的数字资源聚合方法难以满足知识发现与知识利用向智慧层级的更进一步层次的转变,本文在比较现有数字资源聚合方法特点及优劣势的基础上,结合数字资源聚合方法的发展趋向,具体从概念强化、语义强化、关联强化、应用强化及多维强化层面分析了数字资源聚合方法融合的实现机理,并就数字资源聚合方法融合的应用发展进行了归总,提出了多维聚合融合将会是数字资源聚合方法融合的进一步发展方向,为今后数字资源聚合方法融合理论及就馆藏、语义网、标注与关联方面的多维应用研究提供借鉴与依据。

参考文献

- [1] Etzion O. The World Wide Web: Quagmire or gold mine? J. Oren Etzioni Communications of the ACM, 1996(1): 65- 68.
- [2] 姜灵敏. 信息资源聚合与数据挖掘[J]. 广东: 华南理工大学, 2007: 1- 184.
- [3] 邓胜利. 信息聚合服务的发展与演变研究[J]. 情报资料工作, 2012(1): 79- 83.
- [4] 杜 晖. 基于耦合关系的学术信息资源深度聚合研究[J]. 武汉: 武汉大学, 2013: 1- 227.
- [5] 毕 强, 尹长余, 滕广青, 等. 数字资源聚合的理论基础及其方法体系建构[J]. 情报科学, 2015(1): 9- 14, 24.
- [6] 邱均平, 刘国徽. 基于共现关系的学科知识深度聚合研究[J]. 图书馆杂志, 2014(6): 14- 19, 28.
- [7] 贺德方, 曾建勋. 基于语义的馆藏资源深度聚合研究[J]. 中国图书馆学报, 2012(4): 79- 87.
- [8] 张云中. 从整合到聚合: 国内数字资源再组织模式的变革[J]. 数字图书馆论坛, 2014(6): 16- 20.
- [9] 周姗姗. 基于 Folksonomy 模式的数字资源多维度聚合研究[J]. 长春: 吉林大学, 2014: 1- 164.
- [10] 张云中. 本体与自由分类法的融合机理研究[J]. 情报理论与实践, 2012(2): 35- 40.
- [11] 陈开慧. 本体与分众分类的融合模型研究[J]. 图书馆学研究, 2013(5): 73- 77, 19.
- [12] Gruber T. Ontology of folksonomy: A mash up of apples and oranges[J]. International Journal on Semantic Web and Information Systems (IJSWIS), 2007, 3(1): 1- 11.
- [13] Kim H L, Scerri S, Breslin J G, et al. The state of the art in tag ontologies: A semantic model for tagging and folksonomies[C]. International Conference on Dublin Core and Metadata Applications, 2008: 128- 137.
- [14] 张有志, 王 军. 基于 Folksonomy 的本体构建探索[J]. 图书情报工作, 2008, 52(12): 122- 125.
- [15] 熊回香, 廖作芳. 本体在 Folksonomy 中的应用研究[J]. 情报科学, 2010, 28(2): 274- 278.
- [16] 米 佳. 从叙词表到本体的转换研究[J]. 现代情报, 2009(1): 38- 41, 48.
- [17] 赵东霞, 赵新力. 基于政务主题词表的本体构建研究[J]. 现代图书情报技术, 2008(3): 73- 77.
- [18] 王素芳. ontology 与叙词表的融合初探[J]. 大学图书馆学报, 2005(1): 74- 78.
- [19] 李 超. 一种基于主题和分众分类的信息检索优化方法[J]. 情报理论与实践, 2009(10): 108- 110.
- [20] 李 艳. 分众分类与受控词表的集成研究[D]. 太原: 山西大学, 2013: 78- 93.
- [21] 王 雨. 基于社会网络分析的数字图书馆资源聚合研究[D]. 长春: 吉林大学, 2014: 30- 36.
- [22] 陈兰杰, 侯鹏娟. 数字文献资源关联关系揭示方法研究[J]. 图书馆, 2015(2): 41- 45.
- [23] 邱均平, 方国平. 高校图书馆语义化馆藏资源深度聚合模式及其应用研究[J]. 图书馆学研究, 2014(21): 64- 71.
- [24] 孙中秋, 陈晓美, 周珊珊. Folksonomy 与 SNA 资源聚合类比研究[J]. 数字图书馆论坛, 2014(6): 21- 27.
- [25] 杨 萌, 张云中, 徐宝祥. 社会化标注系统资源多维度聚合机理研究[J]. 图书情报工作, 2013(15): 126- 131.
- [26] 岳巧云. 叙词表与 Ontology 的比较研究[J]. 科技情报开发与经济, 2006(24): 86- 87.
- [27] 刘 春, 黄定光. 本体初探[J]. 现代情报, 2008(1): 38- 39.
- [28] Peters I, Stock W G. "Power tags" in information retrieval[J]. Library Hi Tech, 2010, 28(1): 81- 93.

[作者简介] 马鸿佳, 男, 1979年生, 吉林大学管理学院副教授。

李 洁, 女, 1989年生, 吉林大学管理学院硕士研究生。

沈 涌, 女, 1981年生, 吉林大学公共卫生学院讲师。

收稿日期: 2015- 07- 15