

大数据时代知识融合体系架构设计研究*

王曰芬^{1,2}, 岑咏华^{1,2}

(1.南京理工大学经济管理学院, 南京 210094; 2.江苏省社会公共安全科技协同创新中心, 南京 210094)

摘要: 大数据时代不仅改变着研究者科学交流生态和科研创新的模式, 而且对图书情报领域服务提出新的需求和挑战。大数据对知识生态和学科创新模式的重构, 亟需研究人员从根本上探索大数据资源和技术对知识服务和创新的支撑点。在对知识融合、知识生态等相关概念与研究进行系统分析和综合比较的基础上, 面向学科领域研究知识生态重构的挑战, 分析大数据时代知识融合的需求, 归纳总结关联理论与方法, 研究设计知识融合体系架构与构成要素。

关键词: 大数据; 知识融合; 知识生态; 体系架构; 构成要素

中图分类号: G254; TP311.13

DOI: 10.3772/j.issn.1673-2286.2016.10.003

1 引言

在大数据时代如何有效地挖掘和利用大数据的价值, 是知识服务与管理部门必须思考的重大问题。这一问题需要相关研究人员、实践人员以及服务和管理机构重新审视大数据对于知识服务驱动和重构的影响, 并在此基础上规划、支撑和实现知识源的采集收割、辨伪去噪, 及知识的抽取与表示、关联与激活、融合与进化等, 为社会活动、经济活动与科学研究的创新决策提供支持。

以学科创新研究为例, 一方面, 大数据使得科学研究者对知识的获取和使用从过去的“一种混沌”(知识供给匮乏同时有效知识不足)走向今天的“另一种混沌”(知识供给过载但是有效知识不足), 人们迫切希望从大数据中获取系统化的知识支撑。另一方面, 大数据彻底重构传统知识生态, 体现在知识不再以孤立的种群存在于某种生态位上, 而愈来愈呈现出复杂种群系统特征, 噪音和不确定性并未因为知识的多位性而弱化, 反而愈益强化。同时, 在“数据密集型科学”的第四范式研究支持下, 知识之间的泛在“关联”一方面

可为学科创新研究生命周期各阶段的全景知识提供源泉, 激活学科创新工作者的隐性知识; 另一方面作为学科创新研究中“偶遇”“意外”或“异想”知识的源泉, 为学科创新工作灵感提供动力。大数据对知识生态的重构, 改变了学科创新研究的模式。正如近些年学科之间交叉融合不断衍生新学科, 学科间的藩篱已逐渐瓦解, 不同学科和领域间的知识关联和耦合作用愈来愈明显, 科学研究愈来愈呈现出“数据密集”的特征, 基于“大数据”所得的科学结论愈来愈具有发言权, 使得架构在“大科学”和复杂系统观之上的创新研究愈来愈具有突破性意义。

大数据对知识生态和学科创新模式的重构, 亟需从根本上探索大数据资源和技术对知识服务、创新的支撑点。援引《促进大数据发展行动纲要》, 这一支撑点即“对数量巨大、来源分散、格式多样的数据进行采集、存储和关联分析, 从中发现新知识、创造新价值”“利用大数据、云计算等技术, 对各领域知识进行大规模整合”。

因此, 在大数据时代背景下, 以数据信息采集加工组织为业务、以提供知识服务为宗旨的图书馆情报与文

* 本文得到国家自然科学基金“新研究领域科学文献传播网络生长及对传播效果影响研究”(编号: 71373124)、南京理工大学科学研究基金(中央高校基本科研业务费专项资金)项目(编号: 30916011330)资助。

献学领域,如何针对大数据时代知识生态的变化,将新兴的数据科学技术与现有的理论方法相结合,系统深化地研究知识融合的理论方法,并将研究成果应用到实践,将大数据时代碎片化、多维异构的海量数据融合成用户所需的系统化或创新性的知识,并为用户提供知识定制化服务,是图情领域未来发展的必然选择。本文拟在对知识融合内涵、知识融合体系研究进展及知识生态相关研究进行归纳基础上,分析大数据时代图书情报领域面临的挑战与研究热点,进而针对知识融合需求,设计面向知识生态重构的知识融合体系架构,并研究其构成要素。

2 知识融合体系与知识生态的研究进展

2.1 知识融合的概念及其解读

随着信息技术的不断革新,知识呈爆炸式增长,如何有效管理和利用知识成为知识工程与知识管理领域面临的重大课题,这促使知识表示与知识获取等技术

方法得到长足的进步。20世纪90年代中期,知识获取系统已可表示具体某位领域专家的专业知识,但仍无法综合不同专家的专业知识。由此,以Garner等为代表的学者们提出知识融合(knowledge fusion)的概念,旨在针对某些由于观点不同而需要不同的专家协同参与并深度汇聚的问题,以获取更好的问题解决办法^[1]。随后,在计算机科学、信息科学、工程学、数学、管理学等领域纷纷兴起知识融合理论与技术的研究,并逐渐发展成独立的研究领域,越来越多的学者从自己的研究方向对知识融合进行理解和解读,形成不同的研究观点,归纳主要的论述如表1所示。

此外,还有一些研究者基于知识融合提出其他相关概念。如Chen等提出一种非结构化融合理念,认为当现有的知识体系无法提供所需知识时,这种融合就会发生,新的知识产生于知识融合的过程^[7];Kampis等通过对人群发送信息的分布情况来研究协作知识融合,认为传统的知识融合假设信息具有完整性并可永久性获取具有局限性,因此提出一种动态环境下的知识融合,即人们在互动过程中产生的知识融合^[8]。

表 1 关于知识融合的相关概念解读

经典文献	概念界定	研究视角	分析单元
文献[2]	大量的知识资源分布于各组织机构信息系统网络的节点,为开发这些资源,研究人员希望从不同的资源中动态结合知识,称为知识融合。单纯对分布式数据库的检索不能称为知识融合,还需要结合背景知识进行数据实例的解释和利用	建构+实证	系统层面
文献[3]	从管理者的角度分析知识融合和知识创新,提出管理者需要有效地管理组织内部解决问题的过程,而知识融合旨在解决问题,但大多数情况下,知识融合是在人脑中进行的,只有将交流知识的过程呈现出来,才能从对方的知识中得到启发从而进行创新	理论	组织和群体层面
文献[4]	结合Meijer、Preece、Waterson的研究,将知识融合界定为:为补充不足的知识或者获取新知识,将原资源体系中松散耦合的知识进行集成整合,指出知识融合是一种基于不同资源中知识的协同使用方法	建构+实证	系统层面
文献[5]	提出知识融合的基础概念是重构。首先,知识库被转换成相应的关系图,即用中间知识表示;其次,依据该研究提出的结构简介标准以及语义聚类标准将关系图分解为多个子图;最后,利用子图重构融合知识库的元知识	建构+实证	模型层面
文献[6]	知识融合问题被描述为通过数据挖掘而自动生成的知识,与反映专家领域的专业知识相结合、巩固的挑战性问题。问题解决的最终目标是提供精确合理、易于理解并且可实施的模型	建构+实证	模型层面、方法论层面

国内在篇名提及知识融合的研究论文《基于XML的知识融合与知识库组织》发表^[9],而系统研究知识融

合的文献是在2005年发表的浙江大学博士论文《知识融合中若干关键技术》^[10]。经过十多年发展,有关知识

融合的研究主要集中在计算机软件及计算机应用、金属学与金属工艺、图书情报与数字图书馆、联网技术、机械工业等领域。目前,比较认可的知识融合定义是KRAFT工程项目从知识科学角度所做的两类界定:第一类认为知识融合指从众多分布式异构的网络资源中搜索和抽取相关知识,并转换为统一的知识模式,从而为某一领域的问题求解构造有效的知识资源;第二类强调集成的结果是新知识的产生,认为知识融合是一种服务,它通过对来自分布式信息源的多种信息进行转换、集成和合并等处理,产生新的集成化知识对象,同时对相关的信息和知识进行管理^[11]。

综上所述,尽管不同的学者对于知识融合概念理解的侧重点有所区别,但基本形成一定共识。由此,本文总结为:(1)知识融合的研究对象是不同来源的知识及其依附载体;(2)知识融合研究的主要目标是针对某个问题,或某种知识需求,通过一定的技术手段融合互异知识以求解决问题或满足相关需求;(3)知识融合是一个跨学科的研究领域,主要涉及信息科学、计算机科学、图书情报、数学等不同学科的理论和方法;(4)知识融合的结果可以提供用户所需的有效知识,或对于现有知识体系而言的新知识,并且能够给人们带来额外的价值。

2.2 知识融合体系及其研究进展

通过对文献阅读与归纳来看,目前有关知识融合的相关研究主要集中在知识融合体系框架、知识融合实现模式和知识融合应用实践三方面,其中知识融合体系框架是开展工作首要关注的,相关研究分散于各种文献,本文拟从国内与国外两方面加以总结。

在国内研究方面,早期知识融合体系主要服务于产品设计,是基于知识工程与CAD技术相结合构建起来的,知识库中知识的表示主要采取产生式表示法和面向对象表示法^[12]。刘忠途等对基于知识的CAD系统若干关键技术进行研究,为避免烦琐的知识获取过程及降低知识维护成本,提出在传统特征基础上对知识融合体系进行扩展以实现知识融合的机制^[13];景旭等提出基于UG/KDA的广义知识库系统构建知识融合体系^[14]。还有学者利用XML结构化描述语言实现通过知识融合框架对知识进行表示,以构建知识融合的体系架构^[9,15]。

在国外研究方面,Smirnov等针对基于分布式知识

资源进行知识融合任务的系统开发,提出一种多代理架构的组织原则和属性^[4]。研究还给出知识融合体系与运作过程:(1)捕捉知识,从知识源捕捉知识并转换成可以利用的模式;(2)获取知识,从外部资源获取知识;(3)选择知识,从内部资源选择知识(知识基础);(4)生成知识,通过发现或派生方式从现有知识中产生新知识;(5)内化知识,通过保留获取或选择生成的知识而改变系统知识;(6)外化知识,将嵌入系统中的知识输出给环境;(7)知识管理,计划、协调、协作和控制构成知识融合的过程。有学者指出知识融合的对象是不同的知识体,如Lawry等提出一种利用模糊标签将专家知识与人们学习得到知识进行融合的框架^[16];Hu等提出一种基于网页文本的知识融合框架(Web-Page Knowledge Fusion Frameworks),用以合成网页中存在的知识^[17];Dunin-Kępicz等将融合各种分布式异构知识源的多智能体系统框架以及知识融合机制作为重点,并把融合的知识作为命题动态逻辑的Horn子集进行开发^[18]。

2.3 知识生态的概念及其相关研究

1991年,Pór提出知识生态的概念框架,开始设计企业知识生态系统,认为类似于自然界生态系统,在知识管理中也需要考虑知识所处的生态系统^[19]。知识生态是关于学习与加强的状态,在这种状态中知识创造系统的部分与整体彼此关联;知识生态为组织提供一种借助新技术促进合作成员间协作进而取得效益最大化的框架,并认为知识生态实践的核心即知识生态系统^[20]。早期对知识生态的研究主要是设计和培育知识生态系统,研究如何高效地创造知识、集成知识、共享知识和使用知识^[19]。此外,Pór认为知识生态系统是包含信息、洞察力、人和组织能力等多种要素的自组织系统,可被看作“人-知识-技术”三级网络,人构建的网络源源不断地分享、创造知识网络,不同的知识种群通过交互、竞争、演化、进化,以促进知识生态系统的演进^[21]。知识生态系统的理念是通过改善合作进化网络促进知识实体之间交互实现动态演化的一种知识管理途径,有助于提高决策与创新效率^[22]。

研究知识生态的构建与系统应用随后成为国内外学者关注的重点。如Bowonder等提出组合知识管理与知识生态概念的技术管理要点分析框架,并将技术战略概括为通过调整知识搜索、知识构想、知识创造和知识的演变,以满足不断变化的客户需求、竞争威胁和未

来技术轨迹的过程^[23]; Shrivastava认为管理组织智力资本的最佳方式是借助生态环境与生态系统, 提出表征组织知识生态的系统是由在生产性交流与价值创造共生关系中的许多资源、场所、形式与知识主体种类构成的, 并构建用于商业教育与培训的网络化知识生态系统^[24]; Miller提出用于表达核心非正式学习经验的知识生态系统模型, 该模型由非正式学习交互作用关系构成, 如从信息到创造知识的各种关系、各种知识资源的相互支持关系等^[25]; Chau等为了解决康复工程研究与开发中知识翻译周期不完整以及无法有效辅助目标客户等问题, 提出一种知识生态系统, 以搭建从需求到产品的包括知识流、机构库与网络化知识等的康复工程知识翻译过程^[26]; Thomson等从知识生态系统视角开发了基于Web技术的支持林业持续发展的进化系统, 该系统由人、机构、组织、技术与过程构成, 以支撑知识创造、解释、分发、吸收、利用等复杂流程^[27]; 蔺楠等通过构建知识生态系统将生态系统的进化、竞争、共生等动力机制引入组织知识管理, 使个体与群体的知识, 尤其是将隐性知识从深度个人缠绕、团队情节的束缚中解脱, 实现流动、共享与不断更新^[28]; 宫平等对知识生态思想在图书馆知识管理中的应用进行探索, 运用生态系统的层次结构理论描绘图书馆微观知识生态系统是由以馆员为核心的知识群落、图书馆内部因素、图书馆外部因素等要素构成, 并具体分析知识种群的分布结构、知识交流的动态性、知识个体与知识种群的合作、竞争与演化^[29]; 陈静等从知识生态学的角度出发, 阐述数字图书馆知识生态学的表现, 同时对数字图书馆知识生态体系中知识主体、知识客体、知识环境各要素的特性进行分析^[30]; 王硕等对终身学习视野下泛在图书馆知识生态系统的进化过程进行研究^[31]。

从已有的研究发现, 目前知识融合体系仍停留在数据集成层面, 基本能解决信息孤岛问题, 使数据转化为结构化信息, 但在知识深度加工与创新层面还处于探索阶段。同时, 现有的知识生态系统主要基于组织机构知识管理服务需求架构与要素关联研究, 在图书情报领域面向知识服务的知识生态探索仍以理论层面研究为主。而大数据驱动下的知识生态重构, 使知识服务必须从提供型知识服务向分析预测型知识服务转变, 转变过程涉及大数据时代不同层次的知识融合。提供型知识服务侧重于碎片化数据源的有效集成和组织, 将非结构化的知识表示形式结构化、标准化, 在此基础上形成针对某个问题比较完整的知识; 分析预测型知

识服务侧重于显性知识的深度挖掘并将隐性知识显性化, 结合提供型知识服务, 在知识抽取与表示的基础上进行知识融合, 以形成具有深度揭示学科领域知识结构、演化路径与趋势的新知识。

3 大数据时代知识生态重构要素与知识融合关键问题分析

3.1 大数据时代知识生态的重构及其要素分析

通过研究, 本文从两个角度理解与重构知识生态: 一是将知识看作在环境(人、机构、制度等)作用下可进行自组织、学习和进化的Agent。知识生态涉及知识的种群(Distribution)、交互(Interaction)、竞争(Competition)与演化(Evolution)等基本要素, 相同或不同种群和生态位分布的知识通过不断交互、竞争和演化使知识生态系统不断进化。进化得到的新知识不断被加入语义、规则和模型知识库, 而不断更新和迭代的语义、规则和模型知识库又为知识的交互、竞争和演化提供反应因子。二是将知识看作生态链条的能量成分。知识生态主要涉及原始知识提供者、知识生产者、知识服务者以及知识使用者等要素, 在环境(技术、系统、机构、制度等)作用下, 通过知识供需链的价值传递, 不断推进知识本身的创新和进化以及整个生态系统的发展及进化。

从上述两种知识生态视角出发, 如果以学科创新研究为决策场景, 从整体看, 大数据时代知识生态系统要素应包括支撑从大数据-信息-知识-智慧转化的所有主客要素与条件要素及其之间的作用关系。其中, 知识资源特征与知识种群分布状态影响知识生态系统客体要素的结构关系; 知识的交互、竞争和演化, 以及系统中的知识涌现特征影响知识生态系统客体要素的变化关系; 原始知识提供者、知识生产者、知识服务者以及知识使用者之间的“泛在协同”机制影响知识生态系统主体要素的交互模式; 学科创新研究的正式交流与非正式交流模式的交融特征以及大数据时代面向泛在关联的科学研究第四范式等, 影响主体要素的交互结果。

3.2 面向知识生态重构的知识融合需求与需要解决的主要问题

构建知识生态的目的是高效率地创造知识、集成

知识、共享知识和使用知识,对于图书情报领域而言,即通过构建知识生态有效地实现知识服务。大数据使知识的形态、广度、深度、及时性、准确性及效用等都发生变化。在知识服务层面,从知识流动、价值流动、物质流动的角度来看,知识生态系统的功能在大数据时代发生巨大改变。知识流动主要反映系统内部或者系统间不同知识主体之间的知识转移、交换、共享和应用,使知识处于不断融合、增值的自适应动态过程。大数据、云计算、互联网等新兴信息技术改变知识流动的机制,加快知识流动的速度,提升知识融合的期望。知识的价值是知识主客体对知识流动产生利益的期望;价值流指满足这种期望的一系列价值实现活动,是系统各组成部分联系的纽带。大数据使知识价值流动更具有剧变性和网络性,而知识生态系统中物质流动是价值流动的反映,大数据时代放大了这种反映效果。因此,面向知识生态重构的知识融合需求包括功能、业务、数据、流程、技术、平台及相关保障等内容。

知识融合的目标是针对某个问题或某种知识需求,将不同来源的知识及其依附载体,通过一定的技术手段加以融合转化,并将融合后的知识产品提供给用户。从学科创新研究服务来看,如何从“大数据”中提取创新研究所需的“小知识”全景视图,是大数据时代知识融合面临的迫切需求之一。因此,大数据从哪些方面重构知识生态和学科创新模式;面对这种重构,如何通过知识融合的手段来解决学科创新知识服务中的“混沌”问题;大数据驱动的新型知识融合的基本理论和方法有哪些;在此基础上,应构建由哪些要素组成的面向学科创新服务的、由大数据驱动的知识融合体系架构;上述即大数据时代面向知识生态重构的知识融合需要解决的主要问题。

4 面向知识生态重构的知识融合流程与体系架构设计研究

4.1 面向知识生态重构的知识融合流程设计

根据面向知识生态重构的需求与知识融合需要解决的问题,结合学科创新研究的任务情境,基于“数据密集型科学”的第四研究范式,从异构化、碎片化、实时变化的大数据中捕捉和甄取非噪声知识碎片,从知识碎片中提取知识,构建知识元库,将不断更新和迭代的语义、规则和模型知识(以知识库、本体、知识图

谱等形态存在)作为激活因子,通过知识元的生态反应(交互、竞争、演化等)完成知识的再生和创新,在提供“小知识”全景视图的同时,发现“涌现”知识是解决上述问题的基本手段。同时,为区分于现有知识提供与表现方式,本文提出的学科知识全景图谱将由“基础”“比较”“演进”与“创新主体”四个方面构成。其中,基础知识涉及学科领域问题的相关理论(概念、特征、功能、原理)、方法(操作程序)、应用等;比较知识主要涉及学科领域问题的原理、方法、结果、结论、应用等在不同研究或视角上的对比;演进知识主要指学科领域问题的历史、现状、发展、热点、衍生、嬗变、趋势、问题以及对策等知识;创新主体知识涉及个人、研究机构、区域、国家,以及由创新主体构成的社会网络。由学科知识全景图谱、个性化知识融合图谱和检索结果图谱组合形成的一套完整的知识融合结果体系,成为支撑知识生态重构和新型学科创新研究的知识基础。由此,实现和提供上述知识融合的结果体系,需要一套有效的知识融合流程。如图1所示,本文设计将知识生态要素与DIKW(Data-Information-Knowledge-Wisdom)价值链相结合的知识融合流程。

4.2 面向知识生态重构的知识融合的主要理论与方法

针对大数据时代面向知识生态重构的知识融合需求,以面向学科领域科研创新与决策的知识服务为目标,需要梳理知识融合的跨学科理论与方法,以统领和贯穿大数据时代知识融合的体系架构规划与设计以及知识融合的实现模式和服务途径研究,相关理论与方法如表2所示。

4.3 面向知识生态重构的大数据时代知识融合体系架构设计与构成要素

在前述大数据时代知识融合背景与需求分析及大数据时代知识融合的理论方法梳理的基础上,以复杂系统工程理论与思想为指导,本文进一步研究并提出大数据时代知识融合的体系架构,系统全面地分析各相关要素构成,以为大数据时代知识融合平台构建、组织运行、优化升级提供总体规划和实施指南。总体上,该体系架构及其构成如图2所示,包括功能与应用体系、运行与保障体系以及评估与优化体系的三大构成要素。

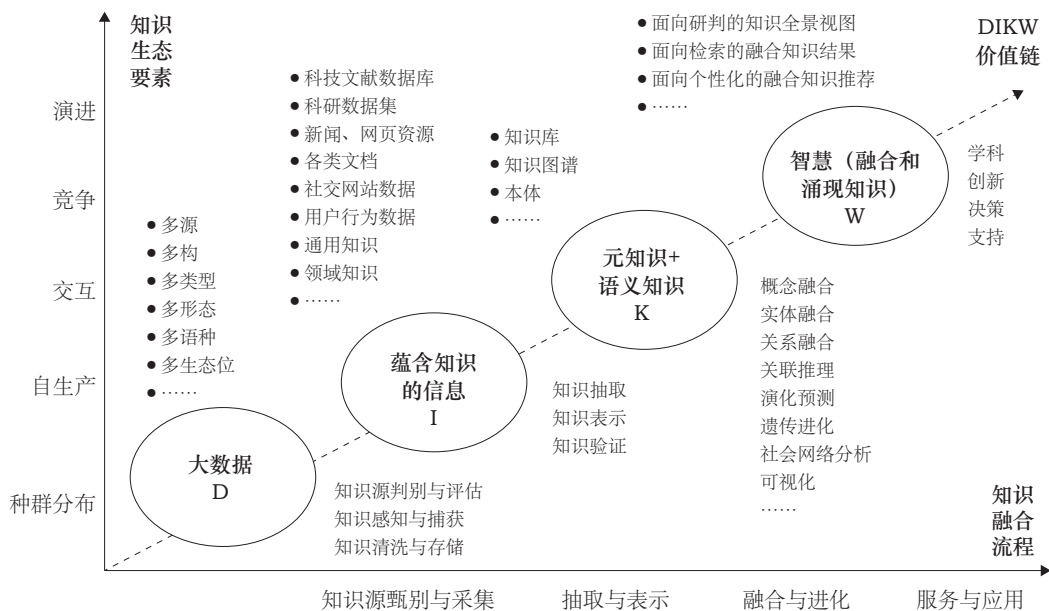


图 1 面向知识生态重构和学科创新服务的知识融合流程——DIKW价值链视角为例

表 2 面向知识生态重构的知识融合的主要理论与方法

理论与方法名称	对知识融合研究的支撑	解决问题层面
知识生态理论	主要涉及生态学领域及图书馆、情报与文献学领域,旨在从生态系统的要素构成与进化机制角度,理解大数据时代面向知识生态重构的知识融合目标、需求与手段,并提供理论依据	目标、需求与保障层面
学科创新决策理论	主要涉及科学学及科学技术与研究管理领域,旨在从学科创新思维、学科创新生命周期过程、学科创新决策支持等角度,为大数据时代面向学科创新研究的知识融合目标和导向提理论和依据	
知识论	主要涉及哲学领域,旨在为课题理清知识的本质、起源和范围提供理论依据	
科学交流理论	主要涉及情报交流与认知心理领域,旨在基于正式交流和非正式交流理论,为泛在知识融合提供思想依据	
用户认知与行为理论	主要涉及认知心理与行为领域,旨在从学科创新研究工作者的认知思维及行为特点角度厘清大数据时代知识融合的内在需求	
复杂系统与复杂网络理论	主要涉及物理学以及复杂系统领域,旨在为分析大数据时代知识生态系统、知识创新与进化、知识网络、社会网络等的复杂性特征提供理论支撑	
信息系统工程理论与方法	主要涉及信息系统工程领域,将大数据时代知识融合平台构建视为一项复杂的信息系统工程,旨在从信息系统工程角度分析大数据时代知识融合的体系架构	实现模式与途径层面
大数据相关理论与方法	主要涉及计算机科学领域,旨在从大数据的感知与捕捉、质量与评估、表示与存储、分析与可视化等层面理解面向学科创新服务的知识融合基础技术问题	
数据挖掘与人工智能理论与方法	主要涉及计算机科学领域,旨在为面向学科创新服务的知识抽取、表示、验证以及融合等提供底层实现途径	
社会网络分析理论与方法	主要涉及复杂网络领域,旨在从知识的社会化特征角度为知识融合提供理论和方法依据	

续表

理论与方法名称	对知识融合研究的支撑	解决问题层面
知识组织、分析与 服务理论与方法	主要涉及图书馆、情报与文献学,从知识服务层面为面向学科创新服务的知识融合提供方法来源和理论依据	实现模式与 途径层面
语义、本体等相关 理论与方法	主要涉及语义学、本体论等领域,旨在为元知识和语义知识的抽取、表示、验证与融合提供理论与方法依据	
可视化与人机交互 理论与方法	主要涉及人机交互和图形、图像领域,旨在从表现层解决大数据时代面向学科创新决策支持的知识融合可视化的基础技术问题	
.....	

第一,大数据时代知识融合的功能与应用体系架构。在大数据时代知识融合的需求和方法体系基础上,以服务于学科创新研究为导向,综合运用各种复杂系统的规划方法以及工程建模方法,提出并设计大数据时代知识融合的功能与应用体系架构,具体包括:大数据资源建设层(涉及对面向学科创新研究的多源异构大数据资源的整体布局与规划、协同建设和有效利用等)、知识源采集与加工层(主要涉及知识源评估与甄选、采集、清洗、转化与存储等)、知识抽

取与表示层(主要涉及从知识源中抽取知识特征项,并对其进行表示和验证,构建知识库、知识图谱和本体库,以形成知识元)、知识融合与进化层(主要通过对接知识库、知识图谱和本体库相关概念、实体及关系等的融合、关联推理、演化分析、遗传进化分析、社会网络分析等手段,为学科创新研究中的学科研判、知识检索、个性化服务等知识需求提供融合的知识视图)以及知识服务与应用(即面向学科创新的决策支持服务)等。

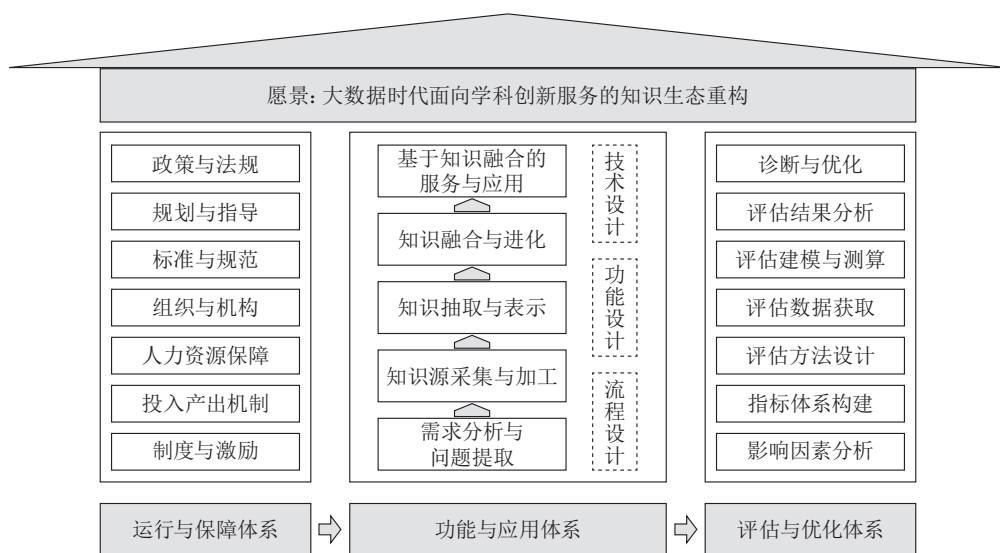


图2 大数据时代知识融合的体系架构设计

第二,大数据时代知识融合的运行与保障体系架构。该结构是支持上述功能与应用体系良性运作的保障体系,包括:政策与法规、规划与指导、标准与规范、组织与机构、人力资源保障、投入产出机制、制度

与激励等。

第三,大数据时代知识融合的评估与优化体系架构。在面向大数据时代学科创新服务的知识融合的影响因素调查与分析的基础上,构建知识融合的评估体

系,形成有关评价指标体系构建,选择评价方法,获取、运算评价数据,分析、诊断与优化评价结果的基本方法论。同时,为保障知识融合流程中各阶段研究结果的可验证性,需要将评估与优化嵌入各阶段,并设计实施。

5 总结

置身于数据爆炸和信息激增的新环境下,面对不断演化的知识生态,人们对于知识的需求,尤其是对经过整理、加工、提炼和聚合的能够支撑社会、经济与科学研究活动的知识需求变得尤为迫切。建立在分布式异构的数据与信息源基础上,为满足信息深度集成、知识扩展开放与知识演化创新需求而发展形成的各类已有知识融合服务,面对大数据时代的到来,不得不重新架构以处理愈来愈碎片化与交错混杂的融合对象和应对愈来愈多样化与深层化的用户需求。现有知识融合服务的框架体系、算法与规则及其应用层次和范围等受到严峻挑战,亟待拓展与创新性的研究。

本文在对知识融合、知识生态等相关概念与研究进行系统分析和综合的基础上,面向学科领域研究中知识生态的重构,分析大数据时代知识融合的需求,归纳总结相关理论与方法,提出知识融合的体系架构及构成。上述研究虽然提出许多观点与结论,但主要还处于理论研究阶段,后续将强化实践应用研究,以为本学科领域的发展提供更多的参考与借鉴。

参考文献

- [1] GARNER B J, LUKOSE D. Knowledge fusion[C]//In Proceedings of the 7th Annual Workshop on Conceptual Structures: Theory and Implementation. Berlin: Springer, 1993: 158-167.
- [2] PREECE A, HUI K, GRAY A, et al. The KRAFT architecture for knowledge fusion and transformation[J]. Knowledge-Based Systems, 2000, 13(2): 113-120.
- [3] MEIJER B R. A management attitude towards knowledge fusion and innovation[C]//Engineering Management Society, 2000, New York: IEEE, 2000: 624-647.
- [4] SMIRNOV A, PASHKIN M, CHILOV N, et al. Multi-agent architecture for knowledge fusion from distributed sources[M]//From Theory to Practice in Multi-Agent System. Berlin: Springer, 2001, 2296: 293-302.
- [5] KUO T T, TSENG S S, LIN Y T. Ontology-based knowledge fusion framework using graph partitioning[C]//Developments in Applied Artificial Intelligence. [s.n.]: Springer, 2003: 11-20.
- [6] MARTENS D, DE BACKER M, HAESSEN R, et al. Ant-based approach to the knowledge fusion problem[M]//Ant Colony Optimization and Swarm Intelligence. Berlin: Springer, 2006, 4150: 84-95.
- [7] CHEN J H, MCQUEEN R J. Knowledge transfer processes for different experience levels of knowledge recipients at an offshore technical support center[J]. Information Technology & People, 2010, 23(1): 54-79.
- [8] KAMPIS G, LUKOWICZ P. Collaborative knowledge fusion by ad-hoc information distribution in crowds[J]. Procedia Computer Science, 2015(51): 542-551.
- [9] 鲍军鹏, 刘晓东, 沈钧毅. 基于XML的知识融合与知识库组织[J]. 计算机工程, 2003(3): 56-57, 138.
- [10] 缪锦. 知识融合中若干关键技术研究[D]. 杭州: 浙江大学, 2005.
- [11] 郭强, 关欣, 曹昕莹, 等. 知识融合理论研究发展与展望[J]. 中国电子科学研究院学报, 2012(3): 252-257.
- [12] 张彦锋, 姜兴渭, 黄文虎. 故障诊断中关联结果与专家知识的融合技术[J]. 哈尔滨工业大学学报, 2002, 34(1): 1-3.
- [13] 刘忠途, 王启付, 陈立平. 三维CAD系统的知识融合与驱动技术研究[J]. 计算机辅助设计与图形学学报, 2005(5): 1013-1018.
- [14] 景旭, 李莉敏, 唐文献. 基于UG/KDA的广义知识库系统的研究与实现[J]. 计算机工程, 2003(4): 124-126.
- [15] 谢能付. 基于语义Web技术的知识融合和同步方法研究[D]. 北京: 中国科学院研究生院, 2006.
- [16] LAWRY J, HALL J W, BOVEY R. Fusion of expert and learnt knowledge in a framework of fuzzy labels[J]. International Journal of Approximate Reasoning, 2004, 36(2): 151-198.
- [17] HU S, CAO Y. Knowledge fusion framework based on Web page texts[J]. Frontiers of Computer Science in China, 2009, 3(4): 457-464.
- [18] DUNIN-KEPLICZ B, NGUYEN L A, SZALAS A. A layered rule-based architecture for approximate knowledge fusion? [J]. Computer Science and Information Systems, 2010, 7(3): 617-642.
- [19] 谢守美. 国外知识生态理论与应用研究综述[J]. 图书情报工作, 2010, 54(18): 103-106.
- [20] PÓR G. Management education and knowledge ecology[EB/OL]. [2016-10-05]. https://www.researchgate.net/publication/265537106_Management_Education_and_Knowledge_Ecology.
- [21] PÓR G. Designing knowledge ecosystems for communities[C]//The Conference on Advancing Organizational Capability Via Knowledge Management, September 29-30, 1997, Los Angeles, USA. [S.l.: s.n.], 1997.
- [22] WIKIMEDIA. Knowledge ecosystem[EB/OL]. [2016-10-05]. https://en.wikipedia.org/wiki/Knowledge_ecosystem.
- [23] BOWONDER B, MIYAKE T. Technology management: a knowledge

- ecology perspective[J].International Journal of Technology Management, 2000,19(19):662-684.
- [24] SHRIVASTAVA P.Knowledge ecology:knowledge ecosystems for business education and training[EB/OL].[2016-10-06].http://www.facstaff.bucknell.edu/shrivast/KnowledgeEcology.html.
- [25] MILLER F.Experiencing information use for early career academics' learning:a knowledge ecosystem model[J].Journal of Documentation,2015, 71(6):1228-1249.
- [26] CHAU T,MOGHIMI S,POPOVIC M R.Knowledge translation in rehabilitation engineering research and development:a knowledge ecosystem framework[J].Archives of Physical Medicine & Rehabilitation, 2013,94(1):9-19.
- [27] THOMSON A J,CALLAN B E,DENNIS J J.A knowledge ecosystem perspective on development of web-based technologies in support of sustainable forestry[J].Computers and Electronics in Agriculture,2007, 59 (1/2):21-30.
- [28] 蔺楠,覃正,汪应络.基于Agent的知识生态系统动力学机制研究[J].科学学,2005,23(3):406-409.
- [29] 高平,徐跃权.图书馆微观知识生态系统的优化研究[J].图书情报工作, 2010,54(21):40-44.
- [30] 陈静,郑建明.知识生态视野下的数字图书馆建设[J].图书馆, 2012(6):39-41.
- [31] 王硕,徐恺英,崔伟,等.终身学习视野下泛在图书馆知识生态系统进化研究[J].图书情报工作, 2012,56(11):23-27.

作者简介

王曰芬,女,1963年生,教授,博士生导师,研究方向:知识服务与数据挖掘等, E-mail: yuefen163@163.com。
岑咏华,男,1979年生,博士,副教授,研究方向:知识挖掘与用户认知等。

Research on Architecture Design of Knowledge Fusion System in the Era of Big Data

WANG YueFen^{1,2}, CEN YongHua^{1,2}

(1.Department of Information Management, School of Economics and Management, Nanjing University of Science & Technology, Nanjing 210094, China;
2.Jiangsu Collaborative Innovation Center of Social Safety Science and Technology, Nanjing 210094, China)

Abstract: The era of big data, it not only changes the scientific communication ecology and innovation mode of the researcher, and it also puts forward new demands and challenges to the service of the library and information fields. Big data to the reconstruction of knowledge ecology and discipline innovation pattern, urgently needs us to explore the support point of big data resources and the technology. Based on the analysis and synthesis of the related concepts and research of knowledge fusion and knowledge ecology, in order to reconstruct the knowledge ecology in the field of study, this paper analyzed the demand of knowledge fusion in the era of big data, summarized the theory and method of knowledge fusion, designed the architecture and components of knowledge fusion system.

Keywords: Big Data; Knowledge Fusion; Knowledge Ecology; System Frame; Components

(收稿日期: 2016-09-27)