



大规模协同知识管理

MASS COLLABORATIVE
KNOWLEDGE MANAGEMENT

朝乐门 著



國家圖書館出版社
National Library of China Publishing House

国家社科基金后期资助项目研究成果

大规模协同知识管理

朝乐门 著

 國家圖書館出版社
National Library of China Publishing House

图书在版编目(CIP)数据

大规模协同知识管理/朝乐门著. --北京:国家图书馆出版社,2014.9
ISBN 978-7-5013-5436-8

I. ①大… II. ①朝… III. ①知识管理—研究 IV. ①G302

中国版本图书馆 CIP 数据核字(2014)第 176663 号

书 名 大规模协同知识管理
著 者 朝乐门 著
责任编辑 高 爽

出 版 国家图书馆出版社(100034 北京市西城区文津街7号)
(原书目文献出版社 北京图书馆出版社)

发 行 010-66114536 66126153 66151313 66175620
66121706(传真),66126156(门市部)

E-mail btsfxb@nlc.gov.cn(邮购)

Website www.nlcpress.com →投稿中心

经 销 新华书店

印 装 北京科信印刷有限公司

版 次 2014年9月第1版 2014年9月第1次印刷

开 本 710×1000(毫米) 1/16

印 张 16

字 数 280千字

书 号 ISBN 978-7-5013-5436-8

定 价 60.00元

国家社科基金后期资助项目

出版说明

后期资助项目是国家社科基金设立的一类重要项目,旨在鼓励广大社科研究者潜心治学,支持基础研究多出优秀成果。它是经过严格评审,从接近完成的科研成果中遴选立项的。为扩大后期资助项目的影响,更好地推动学术发展,促进成果转化,全国哲学社会科学规划办公室按照“统一设计、统一标识、统一版式、形成系列”的总体要求,组织出版国家社科基金后期资助项目成果。

全国哲学社会科学规划办公室

前言

目前,知识管理理论亟待突破性研究。一方面,随着大规模协同生产方式的出现和大数据时代的到来,新兴知识管理实践的不断出现,导致知识管理实践与理论之间的鸿沟日益扩大,使知识管理实践处于无指导状态,加大了知识管理活动的成本和风险。另一方面,知识管理学科的研究责任与研究边界越来越模糊,导致知识管理学科的研究问题无限泛化和研究理论的不断空洞化,使知识管理研究走进泥潭,降低知识管理学科的核心竞争力。

因此,本书从当前知识管理及其相关领域开创性实践的分析入手,以突破知识管理理论研究困境、重新审视知识管理学科的研究责任和核心竞争力、揭示知识管理相关领域新兴实践内在规律以及为未来知识管理实践提供理论指导为研究目的,综合运用知识工程、软件工程、文献研究、调查分析、系统研究和案例分析等科学研究方法,从大规模人机协同的研究视角,依次研究大规模协同知识管理理论的研究背景、目的、基本思想、理论体系、方法论、方法、原则、流程、实现框架及案例分析,为现代组织知识生态系统的建设和知识处理软件系统的研发提供了一种新的知识管理方法——大规模协同知识管理方法,并探讨其实现框架,对于降低组织知识管理成本与风险和解决知识库建设中的知识形式化程度和规模之间的矛盾具有重要参考价值。本书的主要内容如下:

- 第一部分,知识管理研究现状与挑战(第一章)——从知识处理、基于语义 Web 的知识处理、相关理论研究三方面梳理知识管理领域的研究进展,明确大规模协同知识管理的理论基础及进一步研究空间,确定需要重点研究的三个关键内容:提出大规模协同知识管理理论体系、设计大规模协同知识管理方法以及构建大规模协同知识管理方法的实现框架。
- 第二部分,大规模协同知识管理思想和方法(第二至三章)——结合加拿大 Goldcorp 公司的创新、美国 DARPA XC2V 的研发、亚马逊 Amazon Mechanical Turk 的知识处理、卡内基梅隆大学 Receptcha 项目等新兴知识管理相关实践的分析,识别出知识管理实践的十大转变,并提出大规模协同知识管理的理论体系。在此基础上,探讨大规模协同知识管理方法论,并提出大规模协同知识

管理的通用方法、基本原则和主要流程。

- 第三部分,大规模协同知识管理的实现框架(第四章)——以充分发挥人与计算机在知识处理中互补性的优势,为大规模协同知识管理软件系统的研发和大规模协同知识管理知识生态系统的建设提供依据为目的,深入分析人与计算机在知识处理流程中各自的优劣势,借鉴综合集成研讨厅理论,尤其是基于 Internet 的综合集成研讨厅方法论和组织生态系统理论,提出“基于语义 Web 的综合集成研讨厅体系”和“基于语义 Web 的组织知识生态系统建设模型”,并构建“大规模协同知识管理的实现框架”。考虑到大规模协同知识管理方法在组织系统和软件系统中的不同实现方式,在统一顶层设计和保证一致性的前提下,分别提供两个不同的实现框架。
- 第四部分,基于大规模协同知识管理理论的案例分析(第五至第六章)——以大规模协同知识管理方法及其实现框架为解释模型,重点分析两种典型的基于语义 Web 的知识处理项目——KiWi 和 DBPedia,揭示各自的知识处理特征、方法及其框架,分别探讨了存在问题及其解决对策。

大规模协同知识管理是作者首次提出的一种知识管理领域的新理论,主要创新点表现在以下五个方面:

- 提出大规模协同知识管理思想及其理论体系。首次提出大规模协同知识管理理念,并给出其研究背景、研究目的、与现有知识管理理论之间的区别与联系以及新的理论体系。比较有代表性的新观点或(和)新发现有:知识管理与数据管理的区别与联系、知识管理理论体系的火炬模型、知识管理实践的十大转变、知识管理实践与理论之间的鸿沟、大规模协同知识管理理论体系等。
- 提出大规模协同知识管理方法及其实现框架。为了进一步深入分析大规模协同知识管理的本质规律和应用实现问题,本书提出大规模协同知识管理方法论、方法、原则、流程和实现框架。比较有代表性的新观点或(和)新发现有:传统知识管理方法论的困境及其原因、大规模协同知识管理方法论、大规模协同知识管理方法、大规模协同知识管理原则(开放、长尾、专业余、小任务、涌现、自组织、人机协同、工程化处理原则)、大规模协同知识管理流程(知识的表示、联网、发现、获取和集成)以及大规模协同知识管理方法的实现框架(知识处理软件系统的实现框架和组织知识

生态系统的实现框架)。

- 提出知识管理理论体系的主要特征和概念模型——火炬模型。为了分析大规模协同知识管理理论与现有知识管理理论之间的区别与联系,探讨知识管理理论体系的基本特征,并给出知识管理理论体系的概念模型——火炬模型。比较有代表性的新观点或(和)新发现有:知识管理理论体系的五个基本特征(研究对象的层次性、研究范围的广泛性、研究视角的多样性、研究范式的差异性、研究趋势的统一性)、知识管理理论体系的火炬模型和大规模协同知识管理与现有知识管理理论之间的联系与区别。
- 提出知识管理实践的十大转变及现有知识管理研究的局限性。在深入分析加拿大 Goldcorp 公司的创新、美国 DARPA XC2V 的研发、亚马逊 Amazon Mechanical Turk 的知识处理、卡内基梅隆大学 Recaptcha 项目等新兴知识管理相关实践的基础上,提出了知识管理的十大转变和现有知识管理研究的局限性。比较有代表性的新观点或(和)新发现有:知识管理实践的十大转变(从保密到共享、从封闭到开放、从头部到长尾、从专业到专业余、从单一动机到多种动机、从有序到混沌边缘、从小作坊到大工程、从规模增值到利用增值、从安全到信任、从他组织到自组织)、现有知识管理研究的主要局限性(高风险与高成本)等。
- 比较系统地提出语义 Web 与 Web2.0 集成的知识处理模式。从大规模人机协同知识管理视角,以充分利用语义 Web 技术对人机合作完成知识处理任务中的重要作用为目的,在分析基于语义 Web 的知识处理特征和 Web2.0 知识处理方法论的基础上,提出语义 Web 与 Web2.0 集成的知识处理模式和基于语义 Web 的综合集成研讨厅,并选择两个语义 Web 知识管理项目(KiWi 项目和 DBPedia 项目)为案例进行深入分析。比较有代表性的新观点或(和)新发现有:基于语义 Web 的知识处理特征、Web2.0 知识处理方法论、语义 Web 与 Web2.0 集成的知识处理模式、基于语义 Web 的综合集成研讨厅、KiWi 知识处理特征和 DBPedia 项目知识处理特征。

大规模协同知识管理的提出是知识管理理论研究中的重要转折点,突破现有知识管理认识中的基本原则、模式、方法论和方法,重新审视知识管理领域的研究责任和研究本源,分析相关领域中的探索性实践与现有知识管理理论体系之间的鸿沟,实现知识管理理论的继承与创新,对于知识管

理理论研究和实践应用具有里程碑性意义。

- 对于重新审视知识管理领域的研究责任与边界,提升知识管理学科的核心竞争力具有重要意义。大规模协同知识管理理论的提出以大规模协同生产方式和大数据时代为研究背景,结合知识管理相关领域的新兴探索性实践,探讨数据管理与知识管理之间的区别与联系以及知识管理实践的十大转变,重新审视知识管理研究责任和研究本源,提出了一种新的知识管理理论——大规模协同知识管理理论,实现知识管理研究的继承与创新。
- 对于促进知识共享,构建更加开放的知识生态系统具有重要的指导价值。大规模协同知识管理的实现框架将知识链长尾纳入组织知识管理范围,改进了传统知识管理的自上而下的封闭式管理思想,克服目前知识处理模式中计算机知识处理的复杂性和人在知识共享和创新中自我保护性的瓶颈,为建设具有开放性、自组织性、综合集成性的组织知识生态系统提供了统一性、可复用性、半成品性和灵活性的顶层知识管理框架,为语义 Web 环境中的知识生态系统的进一步研究奠定理论基础。值得一提的是,本书为组织知识生态系统的建设和知识处理软件系统的研发提供统一的顶层设计框架,改变知识管理中知识处理软件系统研发与知识生态系统建设之间的脱节现状。
- 对于解决目前知识库,尤其是语义 Web 知识库建设中存在的知识形式化程度与知识规模大小之间的矛盾具有较高的指导意义。目前,语义 Web 知识库建设具有如下特点:一方面,采用非形式化方式表示的知识仍占绝大多数,不利于计算机理解和高效处理,影响计算机知识处理的应用范围。另一方面,采用形式化表示的知识规模不够,不能满足计算机自动处理领域知识和跨领域知识完整性的需要,限制计算机形式化知识处理效果。结果是,语义 Web 知识库要么知识规模足够大,但是其形式化程度不够;要么形式化程度较高,但其知识规模不够大。本书提出的大规模协同知识管理方法及其实现框架对于解决这一矛盾具有重要意义。

本书主要探讨大规模协同知识管理的基本思想、主要原则、理论体系、方法论、方法、实现框架等关键问题。但是,对细节问题的讨论不够,需要尽快解决的两个问题是如何激励大规模协同知识处理的参与者和如何度量参与者及其贡献的可信度。一方面,在基于语义 Web 的知识处理方法

中,主体的参与动机发生变化,主体的被关注度、兴趣爱好、来自其他用户的评价开始成为知识共享的重要动机。因此,大规模协同知识管理中无法沿用传统的激励理论来推动主体参与的积极性。另一方面,本书提出的大规模协同知识管理的参与者具有长尾性,如何度量参与者及其贡献的可信度是为了需要重点研究的科学问题之一。

本书抱着抛砖引玉的宗旨,根据国内外知识管理及相关领域的理论研究和新兴实践的现状与趋势,并结合本人多年研究积累与实践经验,创造性地提出一种新的知识管理理论——大规模协同知识管理。由于大规模协同知识管理理论与实践仍处于起步阶段,相关问题有待进行深层的探讨,加之作者研究水平的限制,对书中的缺陷和问题,恳请读者批评指正。

朝乐门

2014年8月

目 录

前 言	(1)
1 知识管理研究现状与挑战	(1)
1.1 概念界定	(6)
1.2 知识处理	(11)
1.3 基于语义 Web 的知识处理	(28)
1.4 其他相关研究	(60)
1.5 小结	(77)
2 大规模协同知识管理思想	(80)
2.1 知识管理理论体系的火炬模型	(80)
2.2 知识管理实践的十大转变	(86)
2.3 大规模协同知识管理研究目的	(99)
2.4 大规模协同知识管理理论体系	(104)
2.5 大规模协同知识管理的主要贡献	(113)
2.6 小结	(115)
3 大规模协同知识管理方法	(117)
3.1 传统知识管理方法论的困境	(117)
3.2 大规模协同知识管理方法论	(118)
3.3 大规模协同知识管理模式	(122)
3.4 大规模协同知识管理方法	(127)
3.5 大规模协同知识管理原则	(131)
3.6 大规模协同知识管理流程	(133)
3.7 小结	(151)
4 大规模协同知识管理的实现框架	(154)
4.1 构建目的	(154)
4.2 构建方法	(158)
4.3 框架设计	(175)

4.4 小结	(186)
5 案例分析——KiWi	(188)
5.1 KiWi 设计理念	(189)
5.2 KiWi 知识处理模式	(193)
5.3 KiWi 知识处理框架	(198)
5.4 KiWi 知识处理效果评价与改进方案	(202)
5.5 小结	(204)
6 案例分析——DBPedia	(206)
6.1 支持开放、长尾与自组织	(207)
6.2 支持人机协同与专业余	(208)
6.3 支持知识工程与语义 Web	(209)
6.4 支持跨域共享	(213)
6.5 支持增值利用	(215)
6.6 小结	(218)
参考文献	(220)
术语索引	(234)
致谢	(238)

图 目 录

图 1-1	知识管理领域的代表性理论与标准	(18)
图 1-2	语义 Web 分层体系结构	(29)
图 1-3	Web2.0 体系结构	(63)
图 2-1	知识管理理论体系的火炬模型	(81)
图 2-2	美国军用 XC2V 车型	(89)
图 2-3	Amazon MTurk 平台	(91)
图 2-4	Captcha 项目	(92)
图 2-5	ReCaptcha 项目	(93)
图 2-6	知识管理实践的十大转变	(94)
图 2-7	大规模协同知识管理理论体系模型	(104)
图 2-8	大规模协同知识处理	(105)
图 3-1	语义 Web 环境下的综合集成研讨厅体系结构	(126)
图 3-2	大规模协同知识管理的主体交互模型	(130)
图 3-3	大规模协同知识管理流程	(134)
图 4-1	基于 Internet 的综合集成研讨厅的软件体系结构	(164)
图 4-2	基于语义 Web 的综合集成研讨厅框架	(166)
图 4-3	知识生态系统生态圈	(169)
图 4-4	知识生态系统建设框架	(170)
图 4-5	基于 Web2.0 的组织知识生态建设模型	(173)
图 4-6	大规模协同知识管理方法的实现框架	(176)
图 5-1	KiWi 系统的知识处理模式	(189)
图 5-2	以人为中心的知识处理	(191)
图 5-3	KiWi 系统的典型界面	(192)
图 5-4	KiWi 系统中新增知识	(194)
图 5-5	KiWi 系统的知识更新	(194)
图 5-6	KiWi 系统的知识互连	(196)
图 5-7	KiWi 系统的知识导航	(196)
图 5-8	KiWi 系统的知识检索	(197)
图 5-9	KiWi 系统的知识抽取	(198)
图 5-10	KiWi 系统的个性化显示	(199)

图 5-11	KiWi 系统的框架	(200)
图 5-12	KiWi 知识处理效果评价	(203)
图 6-1	DBPedia Live	(209)
图 6-2	DBPedia 实时抽取框架	(210)
图 6-3	DBPedia 本体示意图	(212)
图 6-4	DBPedia 互联数据中枢	(214)
图 6-5	DBPedia Mobile	(217)

表 目 录

表 1-1	语义 Web 的不同定义方法	(7)
表 1-2	人工智能发展的时代划分	(16)
表 1-3	知识管理范式比较	(23)
表 1-4	知识管理模型比较	(23)
表 1-5	第一代知识管理与新一代知识管理的比较	(26)
表 1-6	人工智能和知识管理的比较	(27)
表 1-7	W3C 语义 Web 标准化工作组	(37)
表 1-8	基于语义 Web 的知识处理研究典型项目	(48)
表 1-9	语义 Web 第一代应用与下一代应用的区别	(52)
表 1-10	Web2.0 基本原则	(62)
表 2-1	人机协同知识处理的类型	(106)
表 3-1	语义 Web 与 Web2.0 的比较	(123)
表 3-2	大规模协同知识管理方法与传统知识管理方法的区别	(128)
表 3-3	OWL 中表示知识互联关系的方法	(140)
表 4-1	语义 Web 知识处理中的人与计算机的优劣势比较	(159)
表 4-2	知识生态系统的连续体	(170)
表 4-3	领域语义知识和非领域语义知识的比较	(178)
表 4-4	大规模协同知识管理方法实现框架的特点	(184)
表 6-1	DBPedia 的跨语言摘要	(207)
表 6-2	DBPedia 常用属性	(211)
表 6-3	DBPedia 不同分类方案	(212)
表 6-4	DBPedia 中的 RDF 链接	(215)
表 6-5	DBPedia 的典型应用	(215)

1 知识管理研究现状与挑战

现代社会正进入一个前所未有的大众共同参与经济活动的新时代,大规模协同正在全球范围内改变着发明、生产、销售、分配商品和服务的方式^①。大数据时代和大规模协同处理模式的到来,促使传统知识管理模式发生新变化:

- 知识型员工的自我保护意识加强。随着组织知识管理活动的不断深入,个人的知识保护意识不断加强,在组织知识管理与个人知识管理之间产生博弈关系,使现代组织继续获取“重要员工”的“核心知识”变得越来越难。
- 知识共享的内容和环境发生变化。一方面,人们虽然不愿意共享自己的核心知识,但是非常愿意共享对自己来说不是非常重要的知识。由于核心知识因人而异,当现代组织的知识链足够长时,可以“更小的代价”获取“更重要的知识”。另一方面,人们虽然不愿意在他组织型利益团体内共享知识,但是非常愿意在自组织型非利益团体,如虚拟团队、网络社区共享自己的知识。基于虚拟开放环境的自组织型虚拟组织开始成为未来知识管理的重要场所。
- 用户参与知识管理活动的动机发生变化。个人的被关注度、兴趣爱好、来自其他用户的评价开始成为知识共享的重要动机。
- 大数据时代的到来凸显两种不同主体——人与计算机在知识管理中的各自局限性。数据和知识的几何式增长和分布式存储,更加凸显人和计算机在知识管理中的可互补性优势,人机协同知识管理成为知识管理的必然趋势。

^① Don Tapscott, Anthony D. Williams. Wikinomics: How Mass Collaboration Changes Everything [M]. New York: Penguin Group (USA) Inc. 2006:10-33.

- 对知识管理主体的关注点开始发生变化。众包 (Crowdsourcing)^①、集体智慧 (Collective Intelligence)^②和专业余 (Pro-Am)^③等新理念的提出表明人们对知识管理主体的关注点开始发生新的转变,即从个别“重要员工”转向众多“草根用户”,从“专业人才”转向“专业余”。“专业余”理念认为是“专业余”而不是“专业人才”更适合于跨专业的综合创新,有人甚至提出 20 世纪是“专家”的世纪,21 世纪是“专业余”的世纪^④。

值得一提的是,大数据时代的到来更加凸显知识和数据的区别,要求我们重新审视知识管理的本源、特点和作用。一直以来人们在混淆知识管理与数据管理学科之间的研究责任与研究边界,导致知识管理学科的研究问题无限扩大和研究理论的不断空洞化,使知识管理研究走进泥潭,降低知识管理学科的核心竞争力。大数据时代的到来,不仅导致人们生活、工作、学习和思维方式的大变革,而且将会让我们逐渐认清知识管理与数据管理的区别与联系:

- 数据管理更加重视数据处理的“计算结果”,而知识管理重视知识处理的“涌现效果”。在数据管理中,重视的是数据的简单“量”的计算问题,例如,“1 + 1 等于 2”。但是,知识管理强调知识的“质”的涌现问题。“1 + 1 等于 2”不是知识管理(处理)所关心的问题,而是什么时候“1 + 1 大于 2”、什么时候“1 + 1 小于 2”以及如何在“1 + 1 大于 2”和“1 + 1 小于 2”之间的相互转换才是知识管理真正关注的问题。在知识管理中,“涌现”代表着质变、增值、创新和变革。
- 数据管理更加重视数据的“安全性”,而知识管理则重视知识的“可信度”。知识管理更加强调知识的可信、可共享、可利用,而不是关注如何确保其保密性。在数据管理或信息管理中,安全是指数据或信息的机密性、完整性、可用性、不可否认性、可鉴别性、可审计性和可靠性。但是,在知识管理中可信是指知识主体之间的对等性、知识处理活动的透明性、知识版本之间的可回溯,以及

① Anhai Doan, Raghu Ramakrishnan, Alon Y Halevy. Crowdsourcing Systems on the World-Wide Web [J]. Communications of the ACM, 2011, 54(4): 86 - 96.

② Aaron Weiss. The Power of Collective Intelligence [J]. NetWorker, 2005, 9(3): 16 - 23.

③ Charles Leadbeater, Paul Miller. The Pro-Am Revolution: How Enthusiasts Are Changing Our Economy and Society [OL]. [2013 - 04 - 05]. <http://www.demos.co.uk/publications/pro-ameconomy/>.

④ 张树人,方美琪. Web2.0 与信息系统复杂性变革[M]. 北京:科学出版社,2008:2.

基于行为记录和贡献的可信度。

- 数据,尤其是大数据将成为知识的重要载体。大数据分析不仅可以发现一些人们尚未发现或认知的新知识,而且还可以改变人们的一些常识性知识。例如,大数据时代的到来将会改变人们对数据分析的常识^①,包括数据分析中不再采用随机样本,而是全部数据;不再追求精确性,而是混杂性;不再追求因果分析,而是相关分析等。因此,基于数据管理的知识管理将成为一种新的研究范式 and 方向。
- 相对于数据管理,知识管理更具有复杂性。知识管理中的因果分析更为复杂,重视原因和结果并非是线性关系,小原因可能导致大结果;知识管理中更加强调非称性,任何的一个部分与方面都不能提供充分的信息来实际地或统计地预言其他部分的性质;知识管理更加重视混沌性,而不是完全有序性;知识管理中承认和接受主体的自适应性,认为主体具备了解其所处的环境、预测其变化,并按预定目标采取行动的能力,而且主体的自适应性是影响知识管理的重要影响因素之一;知识管理更加强调层次性,认为不同层次的知识处理任务具有不同的结构和规律,重视跨越不同层次的知识管理与数据管理的区别性。
- 相对于数据管理,知识管理更加强调系统性。一般情况下,数据管理主要考虑数据、应用、服务和平台等底层问题,而在知识管理中,除了考虑上述底层问题之外,更加重视高层问题,包括机制、制度、文化、政策、法律、法规等。因此,知识管理侧重于知识、方法、技术、工具、机制、制度、文化、政策、法律、法规、主体等多个要素的综合集成,而不是片面追求其中的某个(些)要素的控制与优化。

在知识管理研究领域,虽然有一部分专家、学者已觉察到上述变化,但是其研究处于分散状态,尚未进行系统研究。从理论研究角度看,现有的知识管理模式有两种,即以知识管理理论为基础的“人脑知识管理模式”和以人工智能理论为基础的“计算机知识管理模式”。前者侧重于以人为知识管理主体,其研究瓶颈在于如何推动可持续的知识创新和共享;后者以计算机为知识管理的主要主体,其瓶颈在于知识表示的计算机不可理解性。

^① Mayer-Schönberger V, Cukier K. Big data: A Revolution That Will Transform How We Live, Work, and Think[M]. New York: Houghton Mifflin Harcourt, 2013:1-241.