

大数据环境下知识表示与知识组织方法应用

□ 王曰芬 傅柱 / 南京理工大学经济管理学院信息管理系 南京 210094

摘要: 伴随着大数据正在从概念到问题解决的关键转型期,如何发挥核心价值与作用是知识服务业面临的机遇与挑战。论文首先梳理与定量分析知识表示和知识组织理论方法体系及研究进展,其次从实践出发,阐述了在复杂产品设计中的知识需求分析与表示、设计知识的分类与设计领域知识的组织。最后,归纳总结了图书情报领域在知识服务过程中应该实现的三个突破,以及如何抓住四个关键环节的体会。

关键词: 大数据, 知识表示, 知识组织, 复杂产品设计, 实践应用

DOI: 10.3772/j.issn.1673—2286.2014.03.005

1 大数据的发展及其对知识服务业的启示

1.1 大数据的概念与特征

目前,关于大数据的概念还缺乏一个很权威的界定,如: Gartner认为,“大数据”是需要新处理模式才能具有更强的决策力、洞察发现力和流程优化能力的海量、高增长率和多样化的信息资产^[1]; 亚马逊网络服务(AWS)、大数据科学家John Rauser认为,大数据就是任何超过了一台计算机处理能力的庞大数据量^[2]; 维基百科认为,大数据是指无法在一定时间内用常规软件工具对其内容进行抓取、管理和处理的数据集合^[3]。从广义上讲,大数据是从各种类型的海量信息中快速获得有价值信息的能力,人类思维和决策的方式及方法将进入更高层次,有了大数据的这种能力,人类才能真正从“智能”走向“智慧”; EMC公司将大数据工作定义为“数据集,或信息,它的规模、发布、位置在不同的孤岛上,或它的时间线要求客户部署新的架构来捕捉、存储、整合、管理和分析这些信息,以便实现企业价值”^[4]。

尽管大数据的定义呈现出不同的认知与说法,但是所给出的定义基本上都是从其特征出发的。随着研究的深入,关于大数据特征的阐述和归纳已逐渐形成共识,比较有代表性的是3V,即认为大数据具有规模

性(volume)、多样性(variety)和高速性(velocity)的三个特点。因而像Gartner以及大多数的企业就用“3Vs”模型来描述大数据^[3]。除此之外,IDE认为大数据还应当具有价值性(value),大数据的价值往往呈现稀疏性的特点。而IBM认为大数据必然具有真实性(veracity)。所以,有机构在3V之外定义第4个V,真实性(Veracity)为第四特点(也有称第四点为Value,价值密度低)^[5],即“4V”就是Variety、Volume、Velocity及Veracity(或者Value)。目前,业界对以上特点尤其是前3“V”的认识,达成了一致,也有较多的研究成果。

1.2 大数据的起源与发展

1998年,《科学》上刊登的一篇介绍计算机软件HiQ的文章《大数据的处理程序》(A Handler for Big Data)^[1]中第一次使用了大数据(big data)一词。2008年9月《自然》杂志出版“big data”专刊,使“大数据”在学术界得到认可和广泛使用。学者们着手从互联网经济、超级计算、生物医药等多方面关注“大数据”带来的技术挑战、现有技术以及未来的发展方向。2010年,Boillier认为: 计算机存储技术,产生数据流的设备,如摄像机、望远镜和交通监视设备,云计算,面向消费者的应用,如Google Earth和Map Quest,成为大数据产生的几个重要因素,并首次提出“一种新的知识基础设

* 本文系国防科技工业局基础科研重点项目和中央高校基本科研业务费专项资金资助项目(项目编号: 30920130121007)的研究成果之一。

施正在实现,大数据时代正在出现”的观点^[6]。

实际上,早在1980年,著名未来学家托夫勒就在其所著的《第三次浪潮》中热情地将大数据称颂为“第三次浪潮的华彩乐章”。20世纪90年代,“数据仓库之父”的Bill Inmon就开始关注“大数据”。从技术上讲,由于物联网、网络社区、云计算等得到了广泛的应用和发展,才使得大数据在市场上显露出真正的“大价值”。《自然》杂志在2008年9月推出了名为“大数据”的封面专栏,阐述了数据在数学、物理、生物、工程及社会经济等多学科扮演的愈加重要的角色。2012年,大数据概念炙手可热,2013年,大数据走向实践,有的专家称之为“大数据元年”。

另一方面,从Gartner的技术成熟度曲线(Hype Cycle)来看大数据近年来的发展态势,可知大数据的关注度正在急速变化,如图1所示^[7]。

从图1所示曲线可以看出,在2011年大数据作为新技术刚刚上榜,处于技术萌芽期。随着大数据的讨论突破计算机专家范围开始蔓延至产业界和金融界,在

2012年快速处于高速发展期。美国政府的大力推行和引导、各大IT巨头为抢占战略先机和技术制高点而进行的大量投入,使得数据相关的各行业、组织的期望膨胀到巅峰,全民几乎都在讨论这种新技术。2013年,大数据进一步火热发展,新技术的炒作即将达到“过热器”的顶峰,并将由“过热器”转向“低谷期”。

1.3 大数据快速发展带给知识服务业的启示:概念到问题的解决

大数据随着计算方法和信息技术的快速进步而诞生,带来了新的机遇和挑战。面对潮水般涌来的数据,解读数据所包含的信息和知识是当前科学家们致力追求的目标,进而推动了信息技术的发展,带动了大数据的广泛应用,并使其从工具、依据性角色上升到自身蕴含有待开发的信息和知识的新财富地位。大数据已在众多领域成为分析、管理、决策等赖以进行的重要资源,是发现知识、创造知识和认识世界的一种新范式。

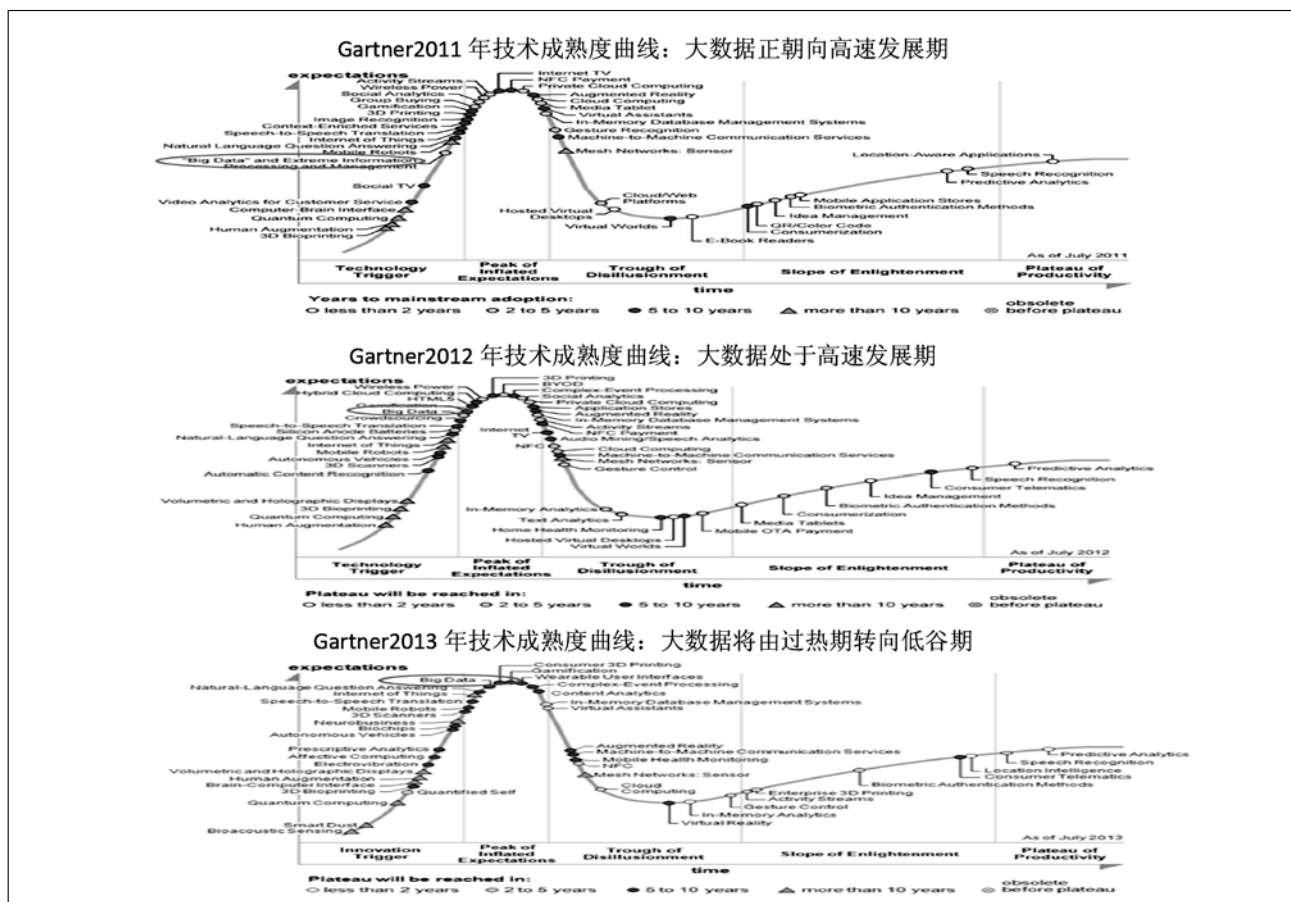


图1 Gartner发布的有关大数据从2011年到2013年的技术成熟度曲线

发达国家纷纷将开发利用大数据作为夺取新一轮竞争制高点的重要抓手,如:2011年英国《自然》杂志曾出版专刊指出,倘若能够更有效地组织和使用大数据,人类将得到更多的机会发挥科学技术对社会发展的巨大推动作用^[8];美国近年来已展开对大数据相关科技项目的大力支持,例如NSF启动了网络化支撑的发现和 innovation (Cyber-enabled Discovery and Innovation) 项目CDI,“从数据到知识”就是其重点研究之一^[9]。大数据将成为描述世界各种物质的新形式、认识事物和表述观点的新方法、预测未来事件的重要依据。对大数据进行挖掘以获取更大的价值,成为下一个创新、竞争、生产力提高的前沿。然而,大数据的运用仍然存在着一些困难与挑战,在数据收集、存储、处理与可视化方面尚有许多问题要解决,尤其在数据处理上,复杂性不仅体现在数据体量大,而且体现在多源异构、多实体和多空间之间的动态交互,用传统的方法描述与度量,无法使数据具有上下文的关联性和导出可理解的内容,更达不到深层次的数据挖掘。

知识服务(Knowledge Services或Knowledge-based Services或KIBS- Knowledge Intensive Business Service)是知识管理、知识组织和知识市场结合而呈现的概念,是一种提供以内容(数据、信息、知识)为基础的组织产出(如建议、方案等),以满足外部用户需要或需求的一种过程^[10]。知识服务的核心是使数据、信息转化为知识,将知识的创新转化为有利于行动的智慧,以支撑生产力和竞争能力的提升。数据作为原始类的产品可经过加工、整理和分析提炼转化为信息和知识,是知识服务业的支撑资源和加工的原材料,对数据进行采集、存储、处理与结果表达构成了知识服务业的价值链。其中,知识表示与知识组织方法是实现知识服务价值链的根本手段,也是体现知识服务业竞争能力的重要基础。面对着大数据的兴起与急速发展,能否把握大数据正在从概念到问题解决的关键转型期,发挥核心价值与作用是知识服务业面临的机遇与挑战。那么,作为知识服务业,如何使大数据的相关研究、技术方法与应用在进入成熟阶段能够有很多本行业主流的产品涌现,并解决知识服务面对的实际问题?在大数据环境下,知识服务的知识表示与知识组织方法广泛而深入发展的路径何在,如何在继承发展中创新?知识表示与知识组织理论方法能否与如何根植在社会实践应用活动中,并成为其他行业领域进行数据处理分析及知识萃取的重要基础与方法支

撑?这些应该成为知识服务在大数据时代着力解决的关键问题。

2 知识表示和知识组织理论方法及进展

2.1 知识表示与知识组织的功能及其主要理论方法体系

(1) 知识表示与知识组织的功能

知识表示(Knowledge Representation)长期以来被认为是人工智能的主要元素,是所有问题解决的关键部分^[11]。其分支领域从认识论中的纯粹哲学方面到解决大规模数据处理中的实际问题,尽管知识表示呈现出不同研究分支的多样性,但是都统一在其解决核心问题上,即以知识可被利用的途径编码各种形式的人类知识。知识表示研究的起点始于Smith B. C.提出的任何一个体现智能化过程的机器都由结构化材料组成的假设^[12],使计算机程序表现一定智能行为而对所需知识采取的结构形式和解释过程是知识表示要实现的功能。知识表示主要承担五种角色^[13]:一种替代物(a surrogate)、一套本体论的约定(a set of ontological commitments)、一种智能推理的片段理论(a fragmentary theory of intelligent reasoning)、一类实用高效的计算方法(a medium for pragmatically efficient computation)和一种人工表达的媒介(a medium of human expression)。知识组织(knowledge organization)最早于1929年由英国著名图书馆学家、分类法专家、《书目分类法》(BC)的编制者布利斯(H. E. Bliss)提出,并在图书馆学、情报学的分类系统和叙词表研究基础上发展起来。目前,关于知识组织的定义主要有以下几种观点^[14]:其一,知识组织是指对事物的本质及事物间的关系进行揭示的有序结构,即知识的序化;其二,知识组织是指对知识客体所进行的诸如整理、加工、揭示、控制等一系列组织化过程及其方法;其三,知识组织是对知识进行整序和提供,既处理大量的现有知识,又能相对降低存储知识的物理载体的盲目增长以免知识过于分散化。所以提供文献、评价科学文献和系统表述以生成新的便于利用和获取的有序化知识单元的处理系统即是知识组织。

知识组织被界定为图书馆与信息科学的名词术语,担负的是文献描述、标引、分类等功能,这些功能主要是由图书馆员、档案馆员、学科专家以及计算机算法来

完成的。随着计算机检索技术等的发展,传统的基于人工的知识组织越来越多地受到挑战^[15]。

(2) 知识表示与知识组织的理论方法

起源于人工智能的知识表示,其任务是将以各种形式存在的知识转化为计算机可识别的编码,然后经过智能化处理后形成有利于人类充分理解与使用的知识。那么,面对人类解决问题的知识需求,如何将知识符号化、形式化及模型化,形成计算机能处理的具有数据结构与系统控制结构的知识,以支持计算机表示和处理知识的可行性与有效性,是知识表示理论与方法研究的核心所在;起源于图书情报领域的知识组织,其任务是将以各种形式存在的混杂的知识,加工处理成具有关联的有序化的可有效存储并可被人类便捷使用的知识。那么,面对知识存储与积累的整序化要求及人类解决问题的知识需求,如何提取知识的特征、建立知识与所表达事物间的映射以及知识之间的关联,形成有序化的知识结构并加以有效存储,以支持知识内容的展示、知识体系的构建、知识检索的实现,尤其是随着科学技术的发展,借助于计算机实现知识组织的自动化与智能化,是知识组织理论与方法研究的核心所在。

知识表示与知识组织虽然起源于不同的学科领域,但是伴随着计算机技术与人工智能技术的发展与普遍应用,知识表示与知识组织不断地走向交融,研究与应用对象扩展至数据、信息、知识层面,研究与应用手段逐渐工具化、自动化与智能化,研究与应用领域扩大到所有需要大规模数据处理、信息加工与知识应用的范围。从知识服务角度看,将数据、信息转化为可被利用的知识,以支持人类的智能化决策的知识需求,是知识表示与知识组织的共同目的。知识表示是知识组织的前提与基础,任何知识组织方法都要建立在知识表示的基础上。

以语言学、逻辑学、统计学、哲学、系统科学、认知科学、脑科学等学科理论为基础,形成解决知识的表示、存储与使用问题的各种模式、策略与方案,具有面向对象、通用化/专业化、推理、分类等特征,构成了知识表示与知识组织的理论方法体系。知识表示的主要方法包括^[16]:状态空间法、问题归约法、谓词逻辑法、产生式规则、语义网络法、框架表示法、面向对象表示法、基于XML的表示法、Ontology表示法等;知识组织的方式主要有两种类型,即以知识单元为基础的知

识组织方式和以知识关联为基础的知识组织方式。知识组织的主要方法包括分类法、主题法、分类主题一体化方法、聚类法、元数据、自动标引技术、自动分类技术、语义网络法、主题地图法、Ontology等。

2.2 基于文献计量的知识表示与知识组织研究的进展

为了比较全面地了解有关知识表示与知识组织研究的国内外现状与发展,本文采用文献计量法,对研究的文献量、所属学科领域与关键词进行了统计分析。中文文献计量分析来源于CNKI数据库中的检索结果,检索策略为关键词“知识表示”或关键词“知识组织”或题名“知识表示”或题名“知识组织”,检索结果为5775条。英文文献计量分析来源于SCIE科学引文索引数据库,检索策略为:主题=("Knowledge representation") OR 主题=("Knowledge organization"),显示结果为5972条。检索时间为2014年1月7日。

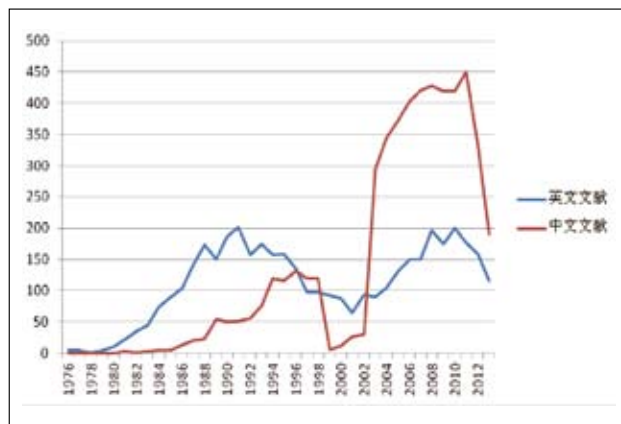


图2 知识表示与知识组织年代分布

(1) 研究概况

由图2可以看出,英文文献中的知识表示和知识组织研究起步较早,1976年英文文献中出现了知识表示或知识组织的词汇,中文文献中直到1981年才开始有这方面的研究。英文文献中的研究从开始一直处于较为稳定的发展状态,到1991年时发文量达到高峰并开始回落,到2001年到达低谷,之后研究持续发展并于2010年再次到达峰顶。中文文献中的研究在发展过程中波动较大,发文量从开始发展到1998都较为平稳,但在1999年时大幅度下降,在之后的四年里有所回升,但

在2003年又出现了大幅度上扬趋势,之后持续稳定上升,并于2011年达到450篇的最高峰。

(2) 学科分布

选取中英文文献研究数量排名前20所在的学科进行统计分析,具体数据如表1所示。

由表1可知,中文文献中的研究绝大多数属于自动化技术、计算机软件及计算机应用和图书情报与数字图书馆领域,约占总篇数的70%,并且这三个学科的文数量和其他学科文数量存在较大的数量差距。这充分说明知识表示与知识组织的研究属于多学科领域,尤其是自动化、计算机领域与图书情报领域的交叉研究占有领先地位。纵观这20个学科类别以及浏览其中的研究内容可以发现,知识表示与知识组织已经跨越起源的研究范围,正在向机械工业、电力工业、金属学与金属工艺、航空等工科领域研究扩展,这也说明知识表示与知识组织在国内各行业的应用正日益广泛;

国外的研究方向主要集中在计算机科学,位列前三位的是计算机科学中的人工智能、信息系统与理论方法子领域,应用主要在电子电气工程、工程制造、运筹管理、通信、工业工程等方面。同时,从每个学科文献数量的统计上看,知识表示与知识组织是跨越多学科子领域交叉研究的。

(3) 研究热点

领域研究热点可以通过相关文献的主要关键词来体现,本文选取出现频次最多的前20个关键词进行统计分析,具体数据如表2所示。

分析表2后可知,关于知识表示与知识组织的研究主要集中在专家系统、决策支持系统和知识库系统构建上,围绕着这三个知识系统的知识工程研究包括用于知识获取的数据结构分析等相关技术,用于知识表示的具有语义的本体模型、Petri网络、形式逻辑、计算语言等数学语言和模型,用于知识应用方面的信息检

表1 知识表示与知识组织学科分布

学科类别 (中)	数量	学科类别 (英)	数量
自动化技术	1422	COMPUTER SCIENCE ARTIFICIAL INTELLIGENCE	3051
计算机软件及计算机应用	1322	COMPUTER SCIENCE INFORMATION SYSTEMS	1502
图书情报与数字图书馆	1030	COMPUTER SCIENCE THEORY METHODS	1278
企业经济	122	ENGINEERING ELECTRICAL ELECTRONIC	923
机械工业	114	COMPUTER SCIENCE INTERDISCIPLINARY APPLICATIONS	907
电力工业	109	COMPUTER SCIENCE SOFTWARE ENGINEERING	677
数学	102	MEDICAL INFORMATICS	300
互联网技术	98	COMPUTER SCIENCE CYBERNETICS	293
新闻与传媒	87	ENGINEERING MANUFACTURING	243
教育理论与教育管理	75	OPERATIONS RESEARCH MANAGEMENT SCIENCE	224
科学研究管理	74	TELECOMMUNICATIONS	217
金属学及金属工艺	63	MATHEMATICS APPLIED	207
航空航天科学与工程	61	COMPUTER SCIENCE HARDWARE ARCHITECTURE	198
工业通用技术及设备	61	INFORMATION SCIENCE LIBRARY SCIENCE	178
武器工业与军事技术	53	ENGINEERING INDUSTRIAL	149
中等教育	48	ENGINEERING MULTIDISCIPLINARY	143
汽车工业	44	ROBOTICS	142
宏观经济管理与可持续发展	41	IMAGING SCIENCE PHOTOGRAPHIC TECHNOLOGY	125
矿业工程	38	ENGINEERING MECHANICAL	114
农业基础科学	36	LOGIC	100

表2 知识表示与知识组织研究热点统计

关键词（中）	数量	关键词（英）	数量
知识表示	1565	Knowledge Representation	2504
专家系统	910	Knowledge Based Systems	596
知识组织	565	Artificial Intelligence	521
本体	380	Expert Systems	459
知识管理	335	Inference Mechanisms	341
知识库	324	Semantics	234
故障诊断	263	Knowledge Engineering	229
知识获取	174	Knowledge Acquisition	210
人工智能	173	Formal Logic	206
图书馆	150	Knowledge Management	178
面向对象	141	Ontologies (Artificial Intelligence)	165
推理机	132	Ontology	165
知识服务	122	Information Retrieval	144
数字图书馆	108	Logic Programming	131
神经网络	102	Petri Nets	127
知识	100	Data Structures	127
推理	74	Mathematical Models	126
XML	73	Problem Solving	125
知识工程	72	Decision Support Systems	125
知识推理	67	Computational Linguistics	123

索、知识推理、神经网络、问题解决和知识管理。其中，基于概念分析与关联建立的Ontology是知识表示与知识组织中可有效解决知识统一表达、智能化处理、知识共享与重用等问题的重要方法。

此外，关于知识表示与知识组织研究领域与应用，国内与国外有一定的差距，国外以计算机领域为主侧重于研究方法、系统开发以及具体应用，如有关的推理机制、语义、形式逻辑、本体、基于知识的系统、专家系统、Petri网、数据结构等。

3 在复杂产品设计中的实践应用

复杂产品设计实践中，需要产品功能、结构、设计方法、设计步骤、图表、公式、材料等各类知识，在设计过程中需要多学科领域人员共同参与和协作，对设计人员要求极高。设计知识的共享和重用已经成为支撑产品创新与开发的关键问题。制造业公司的知识共享

与重用水平非常低，相比服务业公司的42%的知识共享与重用平均水平，而制造业只有28%。由于不仅需要付出额外的努力，而且知识的丢失、不合理应用和错误使知识共享与重用无法有效应用于新产品设计中，所以，在复杂产品设计上，设计知识共享与重用虽然很受关注，但却不易实现。因此，对已有的产品设计、行业标准等通用类知识进行组织存储，发掘专家头脑中的设计经验知识并将其表示出来，以便后续设计的共享与重用，将会对设计者进行产品设计活动有很大的帮助。

然而国内关于产品设计知识表示的实证研究水平总体比较落后，特别在产品设计涉及跨领域的知识时。原因之一是没有形成一个系统的大家公认的知识表示与组织的复用体系，还处于各自为战的研究与应用中。尤其是在涉及多领域、多学科时，由于受到专业知识和各人知识背景的限制，跨领域知识共享与重用相当困难。

那么，从实践出发，如何借助知识表示与知识组织的理论方法，如何使图书情报、知识工程与产品设计等

专家通力协作,实现产品知识的共享与重用,是相关领域急需解决的问题。

下面将以复杂产品设计过程中专家需求的知识表示、产品设计知识的分类、产品设计知识的组织为例,阐述知识表示与知识组织的主要方法——本体(Ontology)的应用。

3.1 复杂产品设计中设计者的需求分析与知识表示

在复杂产品设计中,设计者的需求是知识表示中问题解决的切入点与落脚点,关系到知识表示中采用哪些编码方式、何种数据结构、表达的知识与产品设计过程如何映射以及后续的知识组织等。通过大量的研究与调查,本项目在执行过程中,选取的是专家深度访谈与调查问卷结合的方法。具体地,选取产品设计人员中资历较深的专家,鼓励并引导其提出需求,并对其需求

进行分析与表示。首先,要求产品设计专家以提问的形式表达希望Ontology所描述的知识;其次,对产品设计专家提出的需求作分析,明确Ontology中需描述的知识,可从领域结构知识和领域设计知识两方面分析。经过研究,设计者的需求与分析如表3所示。

3.2 复杂产品设计知识的分类

复杂产品设计中涉及多学科、多领域、多流程的以各种形态存在的知识,那么对复杂产品进行分类,关系到知识表示中编码如何表现、数据结构如何设计以及后续的知识库、知识推理机制如何建立等。因此,依据什么对产品设计知识进行分类,是知识表示与知识组织应用的前提与基础。现有的设计知识的分类一般根据产品设计过程中涉及的知识按照其不同的表现形式进行,缺少统一的分类标准。通过调查国内外产品设计方面的文献资料与访谈专家发现,为了能够完整有效

表3 设计者对本体构建需求的分析

设计者的需求	需要描述的知识
领域结构知识的需求	
我想知道这具体是什么?	定义(对部件的介绍)
我想知道它在整个产品中的作用是什么?	作用(描述部件在整体中实现的功能)
我想知道它是如何实现这个功能的?	原理(描述部件如何实现功能)
具体的工作原理能以图片形式展示吗?	原理图片ID
这个部件的结构位置能以图片展示吗?	结构图片ID
我想知道这个类型有哪些具体型号的实例?	部件实例
具体实例的图片能展示吗?	实例图片
具体型号的实例有哪些具体的属性?	实例的属性
.....	
领域设计知识的需求	
这次设计的总体任务、总体目标是什么?	设计任务(设计需要完成的主要工作和达到的目标)
我想知道参数有哪些取值的限制? 如一个精确的取值推荐, 或一个区间的取值, 或几个离散点之一的限制。	参数取值限制
这种类型部件设计中可供选择的材料有哪些? 选择的理由是什么?	材料选择
可供选择材料的具体属性是什么?	材料的属性
这次设计任务相似的可供参考的设计?	设计实例及设计实例的属性值
已有的设计三维图片实例能展示吗?	设计三维图片ID
.....	

表4 复杂产品设计知识的分类

FBS 模型	功能描述	举例	知识分类
结构 模型	表示产品等设计对象静态结构特征、 属性及属性之间关系	原理、结构、几何等属性特征	描述性 知识
	表示以往设计活动的设计结果及其相关的设计需 求、设计过程等信息	具体部件或以往设计活动的实例	实例性 知识
功能 模型	表示产品设计需要完成的工作和达到的目标，及相 关指导类知识	设计的任务、原则、方法等指导性知识	设计指导 知识
行为 模型	表示事实之间的逻辑推理关系	因果关系、条件关系等	判断性 知识
	表示一个设计过程的操作或操作系列 及其先后关系	任务、操作及其系列	过程性 知识
	表示设计过程中，用数值计算方法 处理的参数等	公式、函数或工具	计算性 知识
	表示设计过程中包含零部件参数的 可选数值或区间	国际/国家/行业/企业标准或手册	手册性 知识

地利用Ontology将产品知识表示与组织出来，必须要用准确、合适的方法描述产品知识模型，而这个模型需要与产品设计流程相切合。项目选择FBS模型（功能-行为-结构模型，FBS由Gero在1990年提出），从功能、行为、结构三方面来描述与分类产品设计知识，通过映像充分反映设计知识的内在联系，即结构支撑行为、行为决定功能，符合产品设计的主旨，并提出了面向FBS模型的产品设计知识分类三原则，即：其一，模块化原则。模型可以分为结构、行为以及功能这三个部分，在表示过程中需要将这三块内容独立表达出来，并在此基础上建立三者的联系，使其语义清晰，关系明确。其二，本体化原则。用各种类型的本体来描述各个模块。采用共享的本体词汇来描述行为与功能、行为与功能分别对应的不同的抽象级别。其三，Web化原则。采用OWL标准化建模语言，充分利用其优点对产品设计知识模型进行表达，以实现产品设计知识的共享，提高设计效率，并进一步实现知识的重用。表4为基于FBS模型的复杂产品设计知识的分类。

3.3 复杂产品设计领域知识的组织——本体架构

知识组织主要的研究对象是存储于文献中的知

识，那么，能否有效地将知识组织方法移植到产品设计过程中，并根据产品设计需求进行继承与发展，是知识表示与知识组织应用的关键问题。本项目采用Ontology方法进行产品设计知识的表达与组织，为更好地实现产品设计领域知识的组织也即本体的构建，需要对获取的知识按本体表达进行组织。按照领域知识本体构建的思想，将产品设计知识按照设计层次进行知识层次结构的区分，具体地划分为结构类知识、设计类知识和图片类知识三个层次。其中，结构类知识是产品设计的基础，设计类知识是对设计的指导，公式图片类知识是对设计过程精确地引导和设计结果的反映。然后，依据知识层次结构对领域知识和设计知识进行组织。所设计的产品设计领域本体的知识表示与知识组织体系如图3所示。

（1）结构知识的组织，包括：1）原理、作用、定义、限制、用途等描述类知识（数据属性，并添加属性值）；2）按不同分类标准进行分类的分类知识（对象间的大类子类关系）；3）子本体的部件组成知识（对象间的部分-整体关系）。

（2）设计知识的组织，包括：1）设计任务、设计目标；2）设计原理、设计指标等设计指导知识；3）设计流程及流程中涉及的参数等计算性知识或手册知识；4）总体重要指标的取值范围或取值限定等经验判断性知

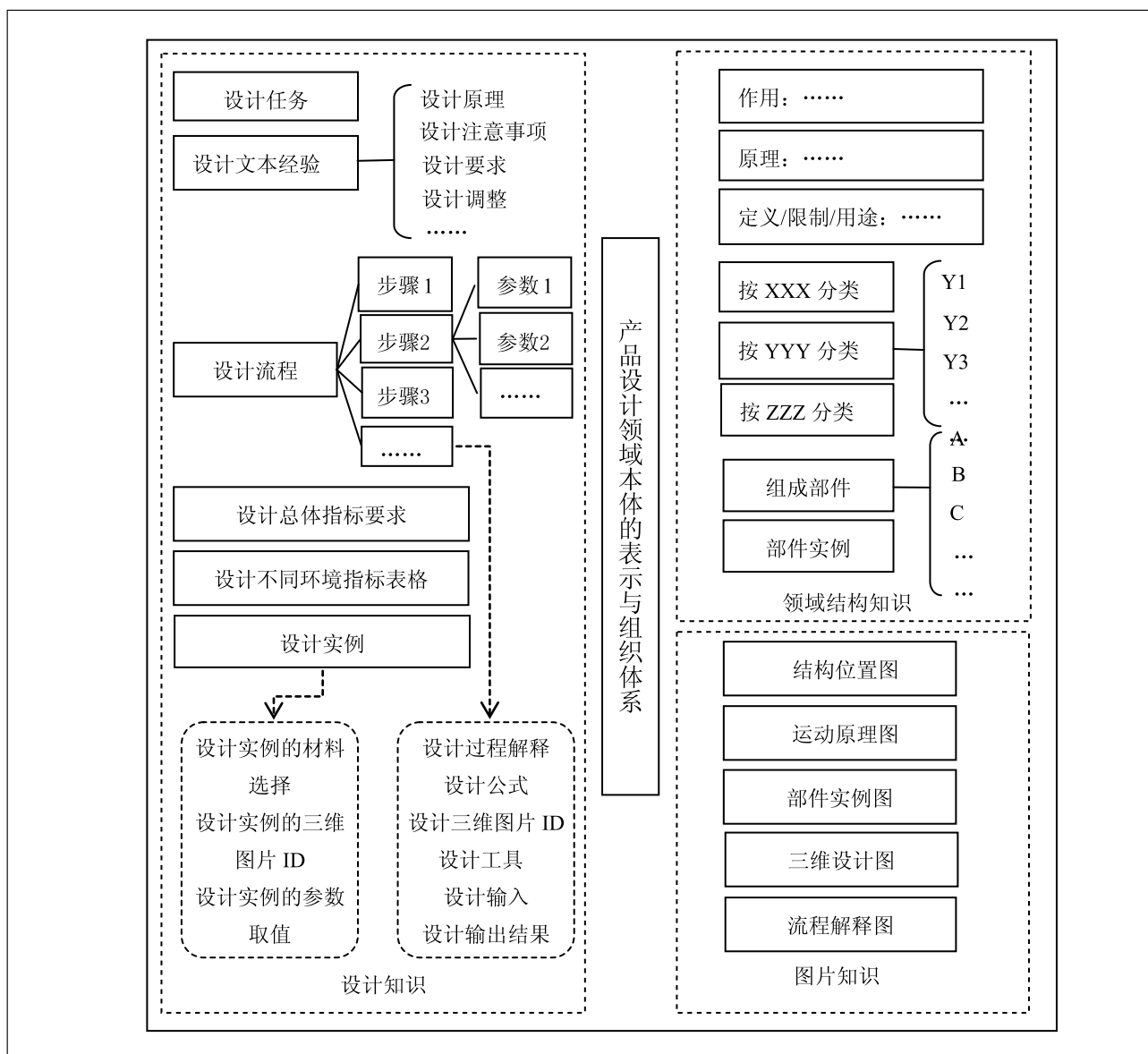


图3 产品设计领域本体的知识表示与知识组织

识；5) 设计流程类和流程中涉及的参数限制等计算性知识；6) 设计实例知识。

(3) 相关图片知识的组织，包括：1) 部件的结构位置关系图；2) 用于说明运动原理的解释图或原理图；3) 部件的具体实例图；4) 设计过程中复杂的设计流程解释图；5) 设计中的三维图。

在上述研究基础上，为方便设计本体的构建，进一步将设计本体细分为总体设计任务、文本性的设计经验、设计公式、设计图片、设计流程及其涉及的参数限制类和具体的设计实例这六块，并对相关概念作如下定义。

定义1: 总体设计任务类知识是指产品设计主要需要完成的工作和达到的目标，是整个设计流程的驱动。因

此构建设计本体的第一步先确定部件设计的总体设计任务和设计任务，从而引导后面的设计指导和设计流程。

定义2: 设计指导类知识是总体设计任务驱动下的具体指导，文本性的设计经验主要包括设计注意事项、设计原则、设计要求、设计调整等能以文本性的形式展现的设计知识。

定义3: 产品设计中的设计图片主要包括三维设计图、部件或整体的实例图、用于说明操作方式的原理图、位置关系的结构图、解释设计流程的流程图五类。

定义4: 设计流程是设计步骤的前后关系的总和，每步流程涉及输入参数、输出参数、设计公式、其他输出结果等要素。

图4和图5是完成的概念本体与产品设计本体的构建实例的可视化展示。

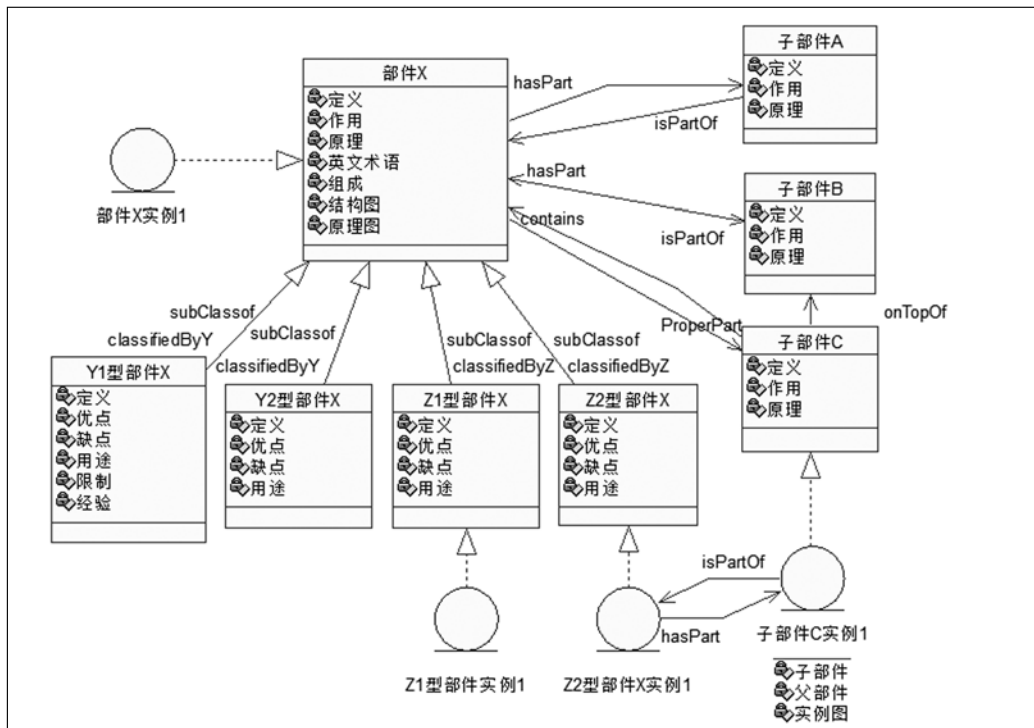


图4 概念本体的建模

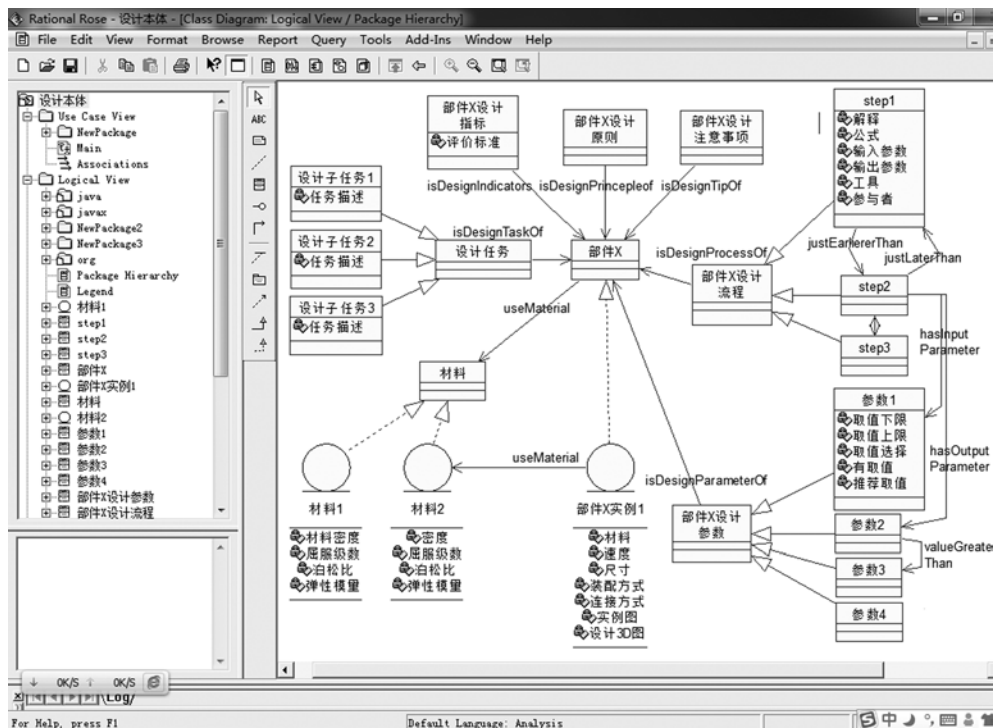


图5 产品设计本体的建模

4 实践研究与应用的体会

信息技术的日新月异,促使知识的生产、创新与利用蓬勃发展。许多行业与机构不仅将知识作为基础资源支撑自身竞争能力与创新能力的培育,而且将知识作为生产要素从事产品开发与服务。大数据的兴起与发展,又推动了这样的进程。作为以知识服务为战略的图书情报领域,如何抓住大数据环境下使数据、信息转化为知识并使其发挥效用的契机,成为赢取未来竞争优势的关键。因此,图书情报领域必须突破服务资源的加工范围,将以文献、数据库为主扩充到文献、数据库、产品资料、实物样本等多种类型;必须突破以目录、摘要、索引、关键词为主的服务内容层次(精度),对表达知识的内外部特征进行不同程度的加工与组织,还需要按照加工程度投入工作人员的隐性知识;必须突破现有的服务模式使服务手段、方式与工具逐渐上升到自动化和智能化水平。那么,如何实现突破成为图书情报领域面临的重大课题。

通过将知识表示与知识组织理论方法应用到产品设计中的课题研究和实践,本文认为图书情报领域在探索突破实现的可行方案和有效路径中,需要重点抓住需求分析、选准面向用途提供服务的模式与途径、规范方法应用的边界与操作标准、保持与专家及用户的互动合作等几个关键环节。

(1) 需求分析一直是图书情报领域开展服务的首要环节,然而在传统的以文献为主的服务产品开发中,以服务机构占有资源的多少与可对数据、信息加工的程度作为需求分析的前提,结果往往导致开发出的产品服务范围受到限制。复杂产品设计不仅涉及多学科、多领域,而且数据、信息与知识的来源广泛,存在方式与表现形式多样化。因此,围绕着产品设计性能参数指标,以设计流程中用户需求为出发点,对产品设计知识的表示与组织需求进行深度分析。如通过对复杂产品设计领域专家的需求分析,可以了解专家对本体所表达的知识和本体需实现的功能的要求。如从领域结构知识需求分析中可以确定本体需要表达部件的定义、作用、原理、结构图、分类标准及具体分类、选择不同部件的原因等知识。

(2) 选准面向用途提供服务的模式与途径,关系到图书情报领域提供知识服务的可行性与有效性。针对知识需求提供深层次的内容加工服务是知识服务的难题,尽管知识表示与知识组织等技术的发展正在逐

渐解决这样的难题。然而,在具体实践中需要利用相对成熟和可操作的技术。因此,在经过需求分析后需要对服务的模式与途径进行选择。例如,在产品设计领域选择提供基本知识的查询、以往设计经验、设计实例的查询或系统推送;在本体构建中选择以表达产品基本属性、特点等结构类知识构成概念本体,以表达产品设计流程、设计规则、设计实例等知识构成设计本体。

(3) 规范方法应用的边界与操作标准,是保障实践应用一致性与持续性的关键。如同分类与标引需要确定操作流程一样,对大量复杂的产品设计知识进行表示与组织不仅工作量大、参与的人数多、耗费时间,而且机器自动化操作基础上还需要人工干预。所以,在技术方法应用前尽可能分析机器操作与专家把关的边界,明确工作的流程,制定与完善每一环节的操作规范。例如制定SUMO顶层本体关系构建的标准、数据属性提取的原则、概念本体与设计本体建模的方法及本体构建的步骤、协同构建本体的操作流程和规范等。

(4) 保持与专家及用户的互动合作,是支撑实践应用的基础。面向复杂产品设计进行知识表示与组织是一项专业性很强的工作,图书情报领域提供的技术方法如果不能与专业人员的知识紧密结合,是难以深入到具体实践环节中的。因此,在知识表示与组织的实践操作中,图书情报领域人员对应用领域专业知识的了解与掌握,以及应用领域专业人员对图书情报领域理论方法的理解与信任,是支撑实践应用的基本前提。这就需要不同领域人员保持持续的对话与互动合作,并坚持研讨与沟通例会制度。

参考文献

- [1] 梁锋.大数据[J].新闻前哨,2013(11):96.
- [2] 畅享网.什么是“大数据”的真正含义?[EB/OL]. [2013-12-10]. <http://www.vsharing.com/k/vertical/2012-5/A658824.html>.
- [3] Big data [EB/OL]. [2013-12-10]. http://en.wikipedia.org/wiki/Big_data.
- [4] 存储时代. EMC紧盯大数据后的大机遇[EB/OL]. [2013-12-15]. <http://stor-age.zdnet.com.cn/stor-age/2011/0223/2010947.shtml>.
- [5] 孟小峰,慈祥.大数据管理:概念、技术与挑战[J].计算机研究与发展,2013,50(1):146-169.
- [6] McKinsey Global Institute. Big data: The next frontier for innovation, competition, and productivity, 2011 [EB/OL]. [2013-12-15]. http://www.mckinsey.com/insights/business_technology/big_data_the_ne

- xt_frontier_for_innovation.
- [7] Gartner2011-2013技术成熟度曲线解读[EB/OL]. [2014-02-10].
<http://www.chinaz.com>.
- [8] 郭贺铨.大数据时代的机遇与挑战[N].求是,2013(4).
- [9] 王飞跃. SciTS: 21世纪科技合作的灯塔? [N].科学时报, 2011-04-13.
- [10] Knowledge market. From Wikipedia, the free encyclopedia [EB/OL]. [2013-10-25]. http://en.wikipedia.org/wiki/Knowledge_market.
- [11] NEWELL A. The Knowledge Level [J]. Artificial Intelligence, 1982(18): 87-127.
- [12] SMITH B C. Prologue to Reflections and Semantics in a Procedural Language [M]. BRACHMAN R, LEVESQUE H J. Readings in Knowledge Representation. Morgan Kaufmann, 1985: 31-40.
- [13] DAVIS R, SHROBE H, SZOLOVITS P. What Is a Knowledge Representation? [J/OL]. AI Magazine, 1993, 14 (1): 17-33 [2013-12-03]. <http://groups.csail.mit.edu/medg/ftp/psz/k-rep.html>.
- [14] 知识组织概述[EB/OL]. [2013-12-03]. <http://wiki.mbalib.com/wiki/%E7%9F%A5%E8%AF%86%E7%BB%84%E7%BB%87>.
- [15] Knowledge organization [EB/OL]. [2013-12-03]. http://en.wikipedia.org/wiki/Knowledge_organization.
- [16] 刘建伟,燕路峰.知识表示方法比较[J].计算机系统应用,2010,20 (3):242-246.

作者简介

王曰芬 (1963-), 女, 博士, 教授, 博士生导师。研究方向: 知识管理、知识挖掘、知识服务。Email: yuefen163@163.com
傅柱 (1990-), 男, 博士生。

Theory and Application of Knowledge Representation and Knowledge Organization in Big Data Environment

Wang Yuefen, Fu Zhu / School of Economics and Management, Nanjing University of Science & Technology, Nanjing, 210094

Abstract: With the key transition from concept to problem solving of Big data, how to play the role is the opportunities and challenges of knowledge services. Firstly, the paper investigated and quantitatively analyzed method system and research progress of knowledge representation and knowledge organization; Secondly, the paper discussed the knowledge requirement analysis and representation, the classification of design knowledge and knowledge organization in complex product design; Finally, it summarized the three breakthrough that should be achieved in the knowledge service process of Library and information field, and the experience of how to grasp the four key links.

Keywords: Big data, Knowledge representation, Knowledge organization, Complex product design, Practice application

(收稿日期: 2014-02-25)