

面向知识服务的知识组织框架体系构建¹⁾

徐绪堪^{1,2} 蒋 勋¹ 苏新宁¹

(1. 南京大学信息管理学院 南京 210039; 2. 河海大学信息管理系 常州 213022)

摘要 在信息爆炸式增长的大数据环境下,由于数据分散、混沌和无序,导致信息爆炸与知识相对匮乏的矛盾日益突出,本文试图通过面向知识服务的知识组织来实现数据知识化、知识有序化以及知识服务化目标,促进传统物理层次的信息组织向认知层次的知识组织的转变。分析国内外学者对知识组织概念、内容以及体系等研究成果基础上,针对用户需求明确知识组织目标、原则和要素,借助粒度原理,从系统角度宏观架构知识组织结构体系,融入用户需求,并通过知识获取与清洗、知识表示与规范、知识挖掘与推理、知识服务与实践等四个阶段实现知识的组织过程,形成一个不断持续改进和循环的知识组织链,使知识服务达到最大的满意度。

关键词 知识服务 知识组织 框架 体系

Knowledge Service-oriented Knowledge Organization Framework and System

Xu Xukan^{1,2}, Jiang Xun¹ and Su Xinning¹

(1. Information Management School, Nanjing University NanJing 210039;
2. Dept. Information Management, HeHai University Changzhou 213022)

Abstract Under the circumstance of big data explosion, the contradictory between information explosion and the relative scarcity of knowledge is increasingly outstanding because of the scattered, chaotic and disorderly data. The paper attempts to realize the goal of data knowledgeable, knowledge ordering and knowledge service based on knowledge service-oriented organization. What's more, we try to promote the transformation from traditional information organizing on physical level to knowledge organization on the cognitive level. Based on the research of domestic and foreign scholars in knowledge organization concept, content and organization system, The goals, principles, and elements of the knowledge organization are designed which are clearly aimed at user's demand. By the principle of granularity, the architecture is built on knowledge organization, the process is consist of four stages which include Knowledge acquisition and cleaning, knowledge representation and specification, knowledge mining and reasoning, knowledge service and Practice, a continuous chain is formed on knowledge organization, then knowledge service is achieved maximum satisfaction through the continuous improvement.

Keywords knowledge service, knowledge organization, frame, system

收稿日期:2013年6月3日

作者简介:徐绪堪,男,1976年生,南京大学博士后,河海大学信息管理系副教授,主要研究方向:知识组织、数据挖掘。蒋勋,男,1980年生,南京大学博士后,主要研究方向:知识库、知识组织。苏新宁(通讯作者),男,1955年生,教育部长江学者特聘教授,主要研究方向:信息处理与检索、知识管理、引文分析等。E-mail: xnsu@nju.edu.cn。

1) 本文系国家自然科学基金项目“面向知识服务的知识组织模型与应用研究”(项目编号:71273126)的研究成果之一。

1 引言

随着社会信息爆炸式增长,信息量极大丰富,形成海量信息或者大数据,给人们工作、学习、研究、日常生活等方方面面带来快捷和方便。同时由于这些信息分散、混沌和无序的状态,导致人类可利用的知识增加不多,用户解决问题感到知识匮乏,信息爆炸与知识相对匮乏的矛盾日益突出,如何有效解决这一矛盾是决定情报学在大数据环境下是否能快速发展的关键。要解决这一矛盾,需要从传统物理层次的信息组织转向认知层次的知识组织,从提供信息服务转向提供知识服务,而对知识组织的研究是这种转变的核心和基础。知识服务对数据的组织有特定的要求,不仅需要通过数据间的联系形成知识,更需要将所有数据在知识服务的框架下构成知识组织体系,知识组织将方便地为用户提供有针对性的知识作为知识组织的目标,通过有效地组织知识实现知识服务。然而,我们目前正处在信息泛滥、信息无序的时代,大量过剩的信息往往使我们束手无策,迫切需要能够将其经过科学处理、整理、关联,最终形成能够解决用户问题的知识。实现这样的目标需要将杂乱的信息蜕变、升华,使之有序、关联和可用,也就是说知识组织必须保证数据知识化、知识有序化、知识服务化。数据知识化是通过分析采集的数据或信息,通过对客观数据加工和处理形成相应的知识,为知识有序化提供基础;知识有序化是对无序的信息或者知识按照一定需求进行分类、关联和管理,实现无序知识有序化,方便知识检索和利用,有利于实现知识服务化;知识服务化是知识组织的目的,最终要让知识利用、共享、传播和增值,根据用户问题环境,融入用户问题到知识组织过程中,让有序化的知识服务于解决用户实际问题。

总之,在知识服务的需求下,在用户需求的驱动下,知识组织理论和方法的指导下,各类技术的支撑下,为实现高效知识服务的目标,知识组织是基础,如何架构知识组织,实现数据知识化、知识有序化以及知识服务化,是我们实现知识服务目标的基础,同时有利于促进图书情报领域从信息科学向知识科学转变。

2 相关文献评述

2.1 国外知识组织研究概述

国外知识组织研究侧重实践方面,主要以

NKOS (Networked Knowledge Organization Systems/Services) 作为主要的研究领域。“知识组织”的概念始见于 1929 年美国图书馆学家布利斯的专著,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的,是信息组织的最高形式。Sue Ellen Wright^[1] 将知识组织系统不仅界定为一组概念的集合,而且还包涵概念之间的语义关系。Zeng Lei^[2] 认为 KOS 可为一个领域内语义结构建模,并为标签、定义、关系和性质提供语义、导航和翻译,嵌入到网络中帮助用户发现和检索知识。Mai^[3] 从认知视角全面分析了人的信息行为对设计受控词汇的影响,将用户来源和价值观作为一个维度进行分析,研究中结合用户需求来描述和组织知识。Park^[4] 等提出以一个以内容为中心、以分面分类法作为设计本体结构的方法,并通过酒本体的实例展示图书馆和信息科学中的知识组织方法。June Abbas^[5] 2011 年出版的专著,从传统知识组织表示结构,知识组织中个人信息的表示,和 Web2.0 下个人和社会信息的表示方法三个方面,综述了知识表示的各种格式和方法。Pastor-Sanchez^[6] 将 SKOS 与其他的词典、叙词表表现方式进行比较,并建议从用户视角来管理、检索等操作,而且提出从用户需求角度研究 SKOS。Smith^[7] 等通过使用传统的知识组织系统原理和其他语义工具,构建了特殊领域的高度结构化知识组织模型,通过知识基础和可视化工具表示科学概念的知识。Bonome, María^[8] 从复杂系统视角分析知识组织系统,认为知识组织系统的新任务是面对复杂多变外界环境和面临可能发生的各种问题,通过知识组织为决策者提供可能的解决办法。Souza Renato Rocha^[9] 等提出一个全新集成的框架对知识组织系统进行分类,并把大量素材和将来用途作为知识组织系统分类的新基础。Donald 等利用 Google Knol 工具,体现同级评审过程中透明性和包容性,通过产生、修改完善、评估和适应以及发布四个阶段进行开放创新知识设计^[10]。美国国家医学图书馆为了提供高效的知識服务,实施了整合信息系统、统一医疗语言系统等项目,实现了医学科研数据、患者健康信息、文献资源等医学信息的有效关联和集成^[11]。

2.2 国内知识组织研究概述

国内知识组织研究成果侧重在知识组织系统概念、基础理论以及体系等方面。1991 年,刘洪波^[12] 指出图书馆内部活动的本质是知识组织,图书馆知

识组织是对人类知识体系结构的模拟和对大脑知识记忆结构的近似模拟,已经在知识组织中考虑知识服务对象个体的因素。1998年,王知津^[13]在《情报学报》上撰文指出知识组织顺应了信息社会和知识社会的发展,梳理知识组织的定义,并对知识组织的目标及其活动领域进行了深入探讨。姜永常^[14]针对CNKI数字图书馆,探讨其依托的信息资源基础和知识管理系统平台,阐述CNKI数字图书馆知识服务的内在机理和外在机制,并探索为不同行业提供相应的知识服务的应用解决方案。王兰成、敖毅、李留英^[15]提出一种多文献数字信息知识组织框架的构建技术,通过使用元-元数据(m-metadata),在知识组织工具和数字信息源元数据之间搭建一个桥梁,以实现数字信息群的知识组织,为多领域多类型信息系统协同发展和文献信息知识集成提供重要基础。张文亮和徐跃权^[16]从微观、中观和宏观三个层次全面系统地理解和解释知识组织的概念和理论体系,通过知识发现、知识描述最终实现知识的表达、共享与创新。贺德方^[17]从知识组织体系的构建和应用角度,对知识组织体系之间的转化、映射以及标准化等方面研究成果进行梳理和总结,认为用户参与和用户使用优先是知识组织的趋势。毕强^[18]分析知识组织系统的相关研究成果并揭示知识组织系统构建从机器可读到机器可理解的发展趋势。薛春香、朱礼军、乔晓东^[19]分析网络环境下知识组织系统的由来和主要特点,剖析了知识组织系统的元数据和术语描述模型,为了更好地引导知识组织系统建设,构建了从功能、内容和结构三方面的知识组织系统评价指标体系。宋培彦等^[20]研究网络环境下知识组织的特征,提出了基于“分类-主题-用户标签”多级映射的结构化概念空间模型,对网络知识百科组织提出知识组织参考。王曰芬等^[21]研究了面向个性化服务的知识组织机制,将用户需求和用户隐性知识纳入其中,形成了个性化服务的知识组织的过程和方法。王知津^[22]等在已经知识组织研究成果基础上,在国内首次全面、系统、深入地探讨网络环境下知识组织的特点和规律、知识组织的新理论和新工具,为网络环境下知识组织实践活动提供指导。白海燕和姜波^[23]围绕知识组织系统的构建与管理、知识组织系统的互操作、知识组织系统的存储与管理、语义元数据的生成系统,对数字图书馆的知识组织体系的构建提出最佳实践和应用说明。董慧等^[24]提出将本体分子应用于数字图书馆领域,建立一个以本体、本体分子为核心的数字图书

馆知识组织四层模型,分析动态知识组织层的实现机制,较好地解决核子、离子和本体分子等知识组织问题。以侯汉清^[25]为首的研究团队对知识组织系统在农史领域应用进行研究。2011年,中国科学技术信息研究所承担了国家“十二五”科技支撑计划项目“面向外科技文献信息的知识组织体系建设与应用示范”研究的相关课题,对叙词表在网络环境下的总体形态、选词方法、词间关系建立、编制管理机制、维护方式、应用方式等方面展开了全面研究,项目立足更大限度地发挥国家科技文献信息资源的作用,满足科技创新对知识服务的需求,面向外科技信息的有效组织与利用。王军和卜书庆^[26]探讨网络环境下知识组织规范的设计,其目标将NKOS和数字馆藏的元数据等知识资源转换为互联网上的关联数据,并成为语义Web的知识关联枢纽。

2.3 基于粒度理论的知识组织研究概述

J. R. Hobbs^[27]提出粒度理论,并从概念与知识的关系角度揭示了知识的粒度特性,说明抽象程度不同的知识之间存在联系,但没有具体给出表示这种联系的方法,所以无法应用于问题解决。卜东波、白硕、李国杰^[28]从信息粒度的角度来剖析聚类 and 分类技术,利用信息粒度原理的框架来统一聚类和分类,在一个统一的粒度下进行聚类计算,而在不同的粒度下进行分类计算。赵昌葆^[29]以工程知识为背景,以问题求解为目的,建立基于粒度知识的综合求解方法,提出概念模型、应用模型与解释模型三次集成的工程知识管理模型,并在工艺设计方面进行实践应用,但知识管理模型主要侧重显性工程知识组织和推理,对于特定工程领域有待融入情景和专家经验知识,跨领域知识集成和运用,解决工程问题的求解有待进一步优化。杨人子和严洪森^[30]针对知识化制造系统自重构中知识网检索方法过于主观以及重复检索和运算等问题,提出基于信息粒度的知识网的模糊分类和检索方法,解决了自重构运算导致的知识网存在多样性的问题,考虑知识网在质、量和复杂性等方面的差异,按照用户需求,将知识网库中的知识网进行聚类,使得最终新知识网的复杂性降低,而且可以得到目标知识网满意度的排序,有效提高分类的正确率,消除聚类结果和先验知识之间的主客观不谐调性。但聚类模型中相似性测度函数的不同影响聚类结果的唯一性,对知识网网的完善度和匹配度的获得科学性有待进一步改善。

通过以上分析可以看出,国外的研究主要从实践视角探讨了信息的本体组织、知识检索、知识粒度化等知识的组织技术,虽有考虑用户需求的知识组织,但缺乏从系统视角研究知识组织框架或模式,也没有从解决用户问题的角度去深刻地探讨知识组织体系及其构成,虽有人利用粒度原理组织知识,但大多集中在词计算、具体工程领域知识的粒度化等方面,缺乏基于用户需求驱动的知识组织过程。国内知识组织主要集中在知识组织的概念和理论研究,主要体现在知识的结构化组织、个性化(服务)组织、系统化组织、集成化组织和可视化组织等方面,由于限于具体技术方法实现等问题,SKOS、OWL表示的研究仍处于实验阶段,还缺乏理论结合实践的研究,虽有考虑用户个性化特征的知识组织成果,主要集中在知识可视化等知识服务平台的个性化服务探索,但从知识服务视角进行知识组织研究成果鲜少,大部分基于知识组织的知识服务平台还处在框架搭建阶段,还在不断探索和完善中,离基于知识服务的应用平台还有很大差距。因此,本文在已有国内外知识组织理论和方法基础上,从知识服务视角对知识组织宏观构架,从系统视角设计知识组织的过程,基于用户需求进行知识的组织和管理,提高知识服务的应变能力、解决问题能力、预测能力、决策支持能力,最终实现数据知识化、知识有序化和知识服务化的目标。

为了实现新环境下知识组织的目标,面向知识服务的知识组织需要宏观架构,明确知识组织的目标和原则,重视知识组织的前期准备环节,明确知识组织包含知识客体、知识提供方、最终用户、知识组织理论与方法以及知识组织环境等要素,对用户需求充分分析基础上,利用各种技术和方法从系统上规划和设计知识组织各个环节,试图探讨知识组织各个环节上知识的粒度变化规律,建立面向知识服务的知识组织体系,从以知识客体为中心向以用户为中心演变,服务从文献服务到信息服务,然后提升到知识服务。

3 架构知识组织宏观结构体系

面向知识服务的知识组织框架应从系统视角出发,以解决用户实际问题为目标,在用户需求驱动下,制定知识组织目标和原则,明确知识组织要素及其体系结构。

3.1 知识组织原则

为了使得知识组织序化过程更加有助于知识服务,在知识组织过程中需要按照一定原则进行构建和整合各种载体,提高各种资源利用效率。面向知识服务的知识组织主要遵循以下原则:

(1)科学性原则。科学性是一切研究工作的前提和基础,也是知识组织的首要原则^[31]。如果不遵循科学性原则,无法有效组织知识,更谈不上为用户提供知识服务,因此只有科学地组织知识,才能保障数据知识化、知识有序化以及知识服务化的实现,结合用户需求进行改进和完善,为知识有序化提供科学支撑。

(2)系统化原则。在知识组织框架体系中,需要系统化地宏观架构知识的组织,并对知识组织各个环节进行系统化组织和协调,全面系统考虑知识组织的整个过程,才能实现知识服务化,同时在知识组织的各个环节能够与用户保持沟通和反馈,各个环节之间进行有效链接,以确保数据知识化和知识有序化,才能提供高效的知識服务。

(3)目的性原则。面向知识服务知识组织是有目的性的活动,即数据知识化、知识有序化和知识服务化,并解决用户实际问题。在知识组织的过程中,对于不同阶段的目的也不一样,在知识组织起初需要获取用户需求,针对不用户不同的需求特征进行知识组织,融入用户不同于专业背景和需求背景,提供有目的性的知识服务。

(4)以用户为中心的原则。以用户为中心大多用于产品或者艺术设计方面尽可能满足用户外在和内在需求,同样地对于面向知识服务的知识组织,分析用户需求和行为基础上,结合当时情景透析用户内在和外在真正的要求,构建用户心智模型,并贯穿在知识组织全过程中,才能为用户提供高效优质的知识服务。

(5)学科性原则。知识组织的最根本的原则是从科学学科中发展起来的,即知识组织的学科性原则让知识更适于知识服务,更有针对性,因为学科不同所需的知识和服务也不同。不论是文献提供还是文献分类,其考虑的核心都是文献含有的知识内容,根据文献的学科特性来组织知识,才能为用户提供有效的知识服务。

(6)共享性原则。由于每个知识组织系统不可能拥有世界上所有的信息资源,共享其他知识组织系统或者数字图书馆的信息资源就很有必要^[32]。

只有通过多个知识组织系统间的协同发展,形成一个互为补充、互为利用、互为推动的文献信息资源保障体系,并且提供网上文献信息服务和知识服务,构建的知识组织系统才能充分发挥作用,是实现知识服务化的基础和保障。

3.2 知识组织构成要素

面向知识服务的知识组织构成要素主要包括知识服务提供方、知识客体、最终用户、知识组织的方法和工具以及知识组织的环境等五部分。在知识组织的环境下,知识服务提供方通过一定方法和技术手段对知识客体进行加工处理,将生产出的知识服务提供给最终用户,知识组织构成要素结构图如图1所示。

知识组织过程一般在一定知识组织环境下进行,借助知识组织方法和工具对知识客体进行分类和处理,对数据进行知识化和知识有序化形成有效知识组织,并提供给知识服务提供方;最终用户结合自身需要提出知识需求,经过知识组织的方法和工具加工和组织后输送给知识服务提供方,知识服务提供方根据用户知识需求针对性组织已有的知识,如果无法满足要求情况下需要从知识客体中重新获取所需知识,利用知识组织的方法和工具形成满足用户知识需求的知识服务,并提供给最终用户,从而实现数据知识化、知识有序化和知识服务化目标。

3.3 知识组织的体系结构

面向知识服务的知识组织过程是问题解决的过程,而求解的问题都是比较复杂的问题,不可能提供现成的解答。在问题解答过程中需要多方面、多层

次的知识,而这些知识具有不同粒度大小,不同粒度知识解决不同层次的问题,因此为了有效组织知识,需要通过一个体系来统一协调和组织,分析知识的粒度在关联和推理过程中粗粒度知识和细粒度知识之间相互转换规律,针对面向知识服务的知识组织过程和特点,将知识组织的体系结构主要分为数据资源层、知识组织层和知识服务层,在进行知识服务过程中每层都是不断完善和更新,如图2所示。

在面向知识服务的知识组织体系结构中,知识资源层是知识组织的基础。该层主要任务完成数据采集,并对数据进行有效处理和有序化组织,其处理和组织主要依据一些数据组织标准,使之有序和可用,也就是说可以在传统的数据库环境实现信息服务。此层的数据主要有:文献数据、事实数据、专业领域数据、知识单元数据以及用户使用行为数据库等,数据的组织方法主要有:MARC格式、元数据格式、一般数据库格式等。

知识组织层主要实现知识有序化和知识再生。知识组织层主要完成知识资源层的数据关联,将数据资源升格为知识资源。在知识组织层将借助一些知识组织工具,建立知识间的语义关系,形成面向特定专题的领域本体、文献链(引用关系)、用户行为知识等知识库,并建立面向决策分析的知识仓库。这些知识库和知识仓库是数据库再组织的知识产品,是实现知识服务的重要保证。知识库和知识仓库将若干个知识点融合形成粒度较粗的知识单元,从而形成用户需求知识、知识单元以及关联知识等不同颗粒度的知识,同时对用户的问题分析和整理后形成问题库,并把用户问题的解答整合形成解答库。

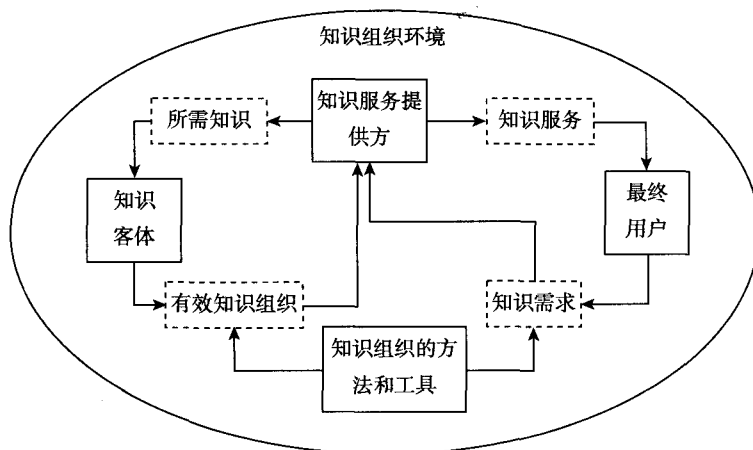


图1 知识组织构成要素结构图

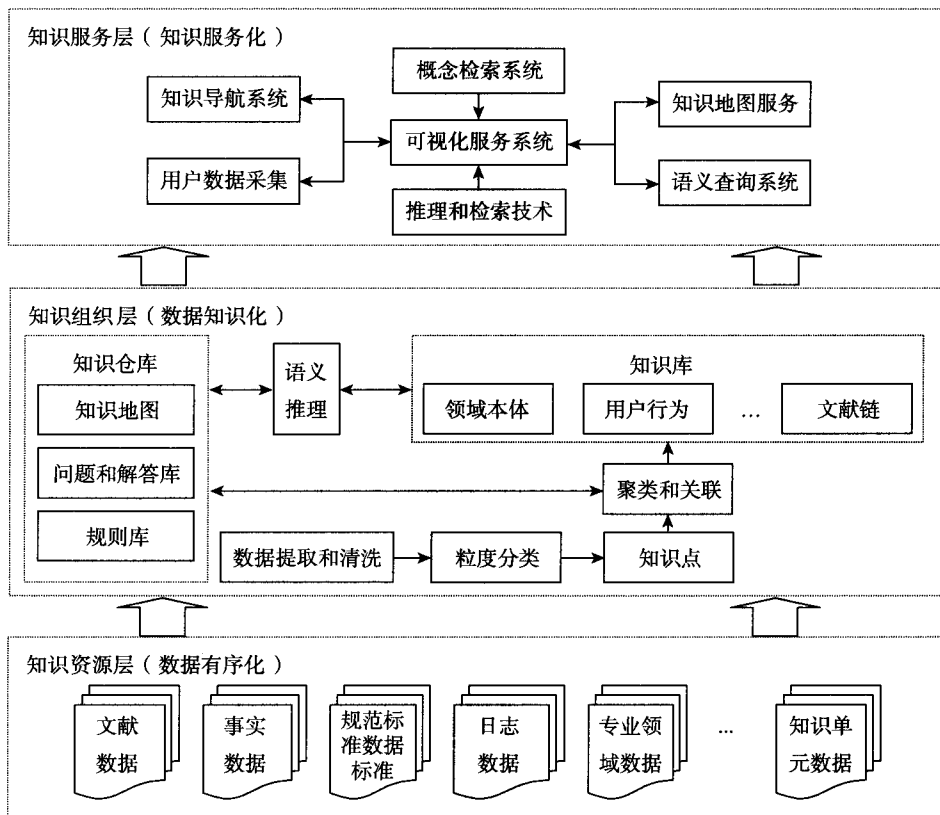


图2 面向知识服务的知识组织体系结构

知识服务层是实现用户需求和知识组织联系的接口层,这一层表现为知识服务平台。该层虽然不承担知识组织的具体任务,但它却是用户和知识组织系统之间的纽带,它一方面根据用户的需求,运用检索技术、推理技术、关联分析技术等对知识库或知识仓库进行运算,从而实现知识服务;另一方面,采集用户知识需求信息和使用系统信息,并将这些信息传递给知识组织层,为建立用户行为分析知识库提供客观数据。在知识服务层中,知识呈现粗粒度动态性,服务客体的知识需求是不确定性和抽象性的粗粒度知识,在知识组织层中经过粗粒度知识细化和细粒度知识重组等环节,形成问题的解答(粗粒度知识),并通过适当方式呈现给服务客体,最终实现知识服务。

目前国内外知识组织体系研究成果集中在用户需求、个性化服务、知识语义组织等方面,中知网CNKI和万方数据虽从宏观、中观和微观等不同层次组织知识,但总体体系结构还是以知识资源来驱动知识组织,虽有从系统角度组织知识的,但构建的知识组织体系难以针对用户具体问题动态重新组织知识并提供高效知识服务。本文构建的知识组织的体系结构以用户需求为驱动来动态地组织知识,在数

据有序化、数据知识化以及知识服务化全过程中,不仅仅从系统角度集成已有知识组织方法和技术,而且还充分考虑人文社会因素,针对用户要解决的问题,借助粒度原理灵活动态组织不同层次和领域的知识,易于实现知识组织实践,为用户提供高效知识服务奠定坚实基础。

4 面向知识服务的知识组织过程构建

为了满足新形势下用户对知识组织的需求,提高知识服务质量,有必要分析传统知识组织方式存在问题和不足。首先传统知识组织中采用统一知识粒度大小进行分类和聚类,但由于解答问题中需要不同粗细粒度的知识进行支撑,所以用户无法快速获取个性化问题的满意解答;其次各类知识的语义关联不足,导致对用户需求针对性不强,检索结果满意度低。显然传统知识组织无法满足用户日益复杂和全面的需求,同时急需以用户为导向的知识组织来提供高效的知识服务。

面向知识服务的知识组织是以有效解决用户问题为目的,因此在问题解答过程中,有时需要细粒度知识;有时需要粗粒度的知识。因此为了有效解决

知识组织各阶段对知识认知规律不确定的问题,从系统角度组织各阶段的知识,为了更好地实现数据知识化、知识有序化以及知识服务化,将知识组织过程可以分为知识获取与清洗、知识表示与规范、知识挖掘与推理、知识服务与实践等阶段。每个阶段对应数据级、知识级、决策级以及服务级等不同粒度的知识,形成从知识获取到知识服务的知识粒度层次,构成一个反馈和动态的系统。其中知识获取与清洗和知识表示与规范阶段主要实现数据知识化,知识挖掘与推理主要实现知识有序化,知识服务与实践主要实现知识服务化。面向知识服务的知识组织过程如图3所示。在知识组织过程中,结合用户需求和知识组织情景,每个阶段都形成相应的粒度知识群,通过知识粒度模型对数据资源物理层次和认知层次知识化和有序化,形成解答问题的新知识,最终实现组织组织的目标。例如,在销售啤酒和尿不湿的超市中,对于数据级粒度的“啤酒及销售量”和“尿不湿及销售量”等知识获取与清洗,形成啤酒销售规律和尿不湿销售规律等知识级粒度知识,然后通过知识挖掘与推理挖掘啤酒和尿不湿销售之间关联,发现尿不湿的顾客同时也购买啤酒等潜在规律,结合这个规律为超市提供知识服务,并通过把啤酒和尿不湿相邻摆放实现知识价值增值。

4.1 知识获取与清洗

虽然国内外学者对知识获取进行了大量研究,但它仍然是知识组织和管理领域的主要瓶颈问题。目前知识获取与清洗的方法主要有模糊理论、粗糙集、神经网络、决策树等机器学习和自然语言分析两大类,侧重于从大量已有数据中挖掘出潜在的联系、

规则或规律^[33],而面向知识服务的知识获取与清洗不仅要挖掘已有数据的规律,还要挖掘用户潜在需求和情境,并基于用户需求进行知识获取。

知识获取与清洗是知识组织过程的输入部分,该阶段主要功能是根据用户的任务首先从各类数据资源中获取所需要的数据或知识,然后对这些数据或知识按照一定要求和粒度原理等方法进行清洗操作,便于知识的表示和应用。该阶段是知识组织过程中核心部分,能否获取丰富、有效的知识直接影响知识服务的质量。

知识获取与清洗的对象主要包括用户需求和各类数据资源两大类。其中用户需求主要是用户知识背景和使用行为、需要解决问题和具体要求等,各类数据资源包括文献资源和热点数据等。

知识获取与清洗阶段抽取步骤主要包括筛选目标资源、知识抽取模式、知识元抽取以及抽取后知识的组织等,其中目标资源筛选主要从海量数据资源中选择最合适的知识来源,排除不必要的数据干扰;知识抽取模式是按照一定规则对目标资源进行人工或者自动知识抽取;知识点抽取主要根据知识抽取模式制定规则提取相应各种类型的知识点;抽取后组织是对抽取的知识点进行存储和组织,如果符合要求直接输送到知识表示与规范阶段,否则返回知识抽取模式步骤,重新完善和调整抽取模式,直到满足要求为止。

4.2 知识表示与规范

知识表示与规范是针对不同类别和不同粒度大小的知识以及知识关联,并按照一定规范表示以便人们识别和理解知识^[34]。知识表示是知识组织的

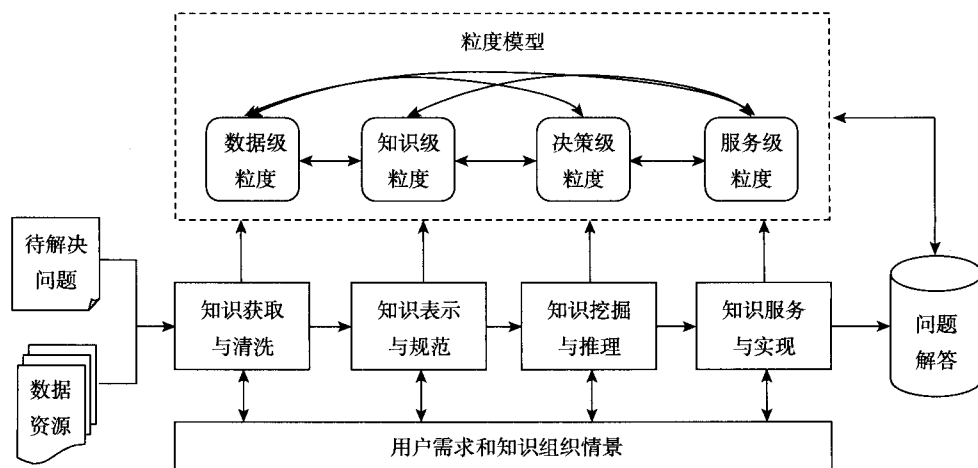


图3 面向知识服务的知识组织过程

前提和基础,任何知识组织方法都是要建立在知识表示的基础上。

知识表示与规范是指在遵循知识组织构建规范或者标准基础上,选择适当的编码方式,结合粒度原理,形成不同层次不同大小的知识粒度,包括用户需要解决的问题、解决问题的中间过程、问题最终解答以及可能涉及到相关知识等。知识表示主要目标是从功能、内容和结构^[19]等方面都能满足解决用户需求、用户和知识服务者沟通以及方便机器理解,以便在知识挖掘、知识仓库构建、知识推理、知识网络知识服务等环节中能够更好地传播和应用知识,是知识组织中共享和互操作的基础。

知识表示与规范对象主要是不同类型和不同粒度大小的知识,包括主观知识和客观知识两种,为了有效地表示和规范知识,应尽可能将事实、数据、信息等知识之间的联系表达或者标注出来。

知识的表示就是对知识的一种描述,或者说是对知识的一组约定,一种计算机可以接受的用于描述知识的数据结构,是知识形式化的过程,是知识结构及其处理机制的综合,即知识表示 = 知识结构 + 处理机制,将知识元作为知识表示的最小单位^[35],其粒度最细。常见的知识表示方法有产生式规则、语义网、框架法、面向对象表示法、过程表示法、剧本表示法等。知识表示主要包括规范化、概念化以及

表示实现三部分。

4.3 知识挖掘与推理

知识挖掘与推理是知识组织过程中核心环节,决定着知识服务效果,主要任务是从数据集中识别出有效的、新颖的、潜在有用的,以及最终可理解的模式的非平凡过程^[36]。在知识表示与规范化成粒度知识后,无法为用户解决实际问题,有必要对粒度知识分类、聚类以及关联等操作,形成静态和动态粒度知识网络,为提供高效知识服务提供有力支撑。

知识挖掘与推理是在知识表示与规范基础上,融合用户需求和情景,以解决用户问题为主线展开知识挖掘和推理。在求解用户问题时,可能根据不同的目的变换问题的粒度大小,对粒度知识进行分类、聚类、引用和关联等操作,寻求隐含知识中的潜在模式或者规律,为问题求解提供不同层次和不同类别粒度知识,通过 Agent 技术根据已解决的问题和知识构建问题和知识 Agent^[37],通过自适应调节机制在均匀、统一粒度下进行粒度知识聚类,在非均匀粒度下进行分类,通过粒度知识之间关联和引用分析,形成一个多 Agent 动态优化学习系统。系统从解决问题出发,通过各 Agent 组之间的交互操作,利用与共同的对象模型之间的数据接口实现相关数据的传递,从而解决用户遇到的不同问题。知

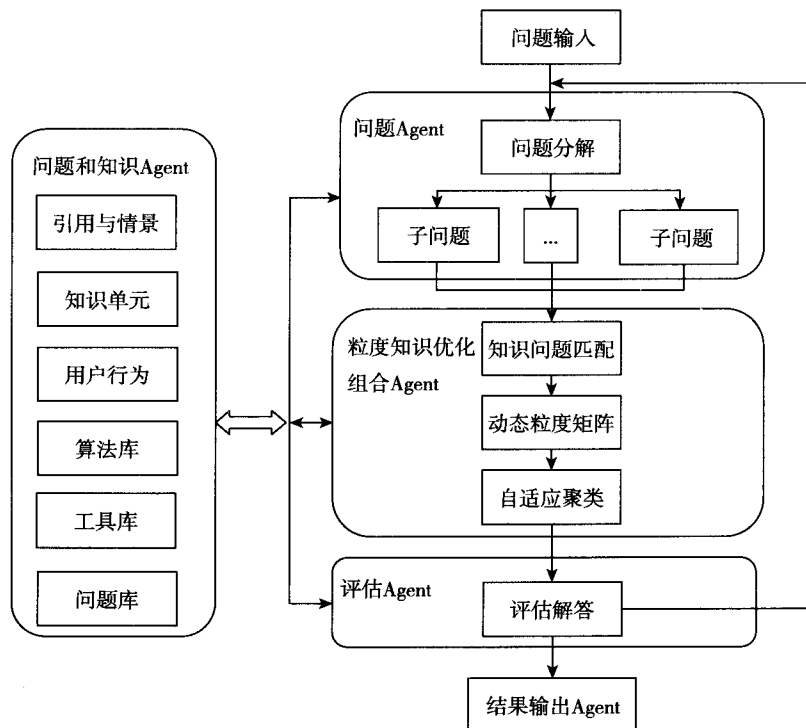


图4 知识挖掘与推理过程

识挖掘与推理过程如图4所示,主要通过知识Agent、问题Agent、粒度知识有序化Agent以及评估Agent对知识进行分类、聚类、推理和挖掘等有序化。

4.4 知识服务与实现

知识服务与实现是知识组织过程的实施阶段,是与用户交互的接口,也是高效知识服务的体现和保障。知识服务与实现侧重向用户提供知识及问题解决方案,不仅仅是简单的信息集合,以知识的搜集、组织、分析及重组信息和知识的能力为基础,根据用户的问题和环境,融入用户解决问题的过程中,借助信息系统设计思路,利用本体、语义网、数据库等计算机信息技术,提供能够有效支持知识应用和知识创新的服务。

知识服务与实现的最终目的是为用户提供所需的个性化知识资源和智力服务,促进用户任务的完成和问题的解决。为了达到这一目的,知识服务和实现的组织是面向用户需求和问题,整个知识服务和实现首先要收集问题解决的情景,形成情境感知的知识需求,并贯穿到整个问题解决的服务过程中。知识服务与实现是一种个性化服务,个性化服务主要包括主题化服务和个人化服务,其中主题化服务是针对特定问题,按照主题来组织和实现服务;个人化服务是针对特定用户的个性化要求提供知识服务。知识服务与实现是一个价值增值过程,为用户提供恰到好处的知识、智力服务和问题解决方案,在提升知识工作者知识获取能力、利用能力和知识创新能力之外,同时也实现了知识服务的价值。

知识服务与实现结合系统设计思路和流程,知识服务与实现可以事先进行知识组织的准备,通过数据接口对各类知识来源库进行知识获取与清洗,过滤不需要的垃圾,按照知识表示的规范对知识进行粒度化表示形成知识仓库,根据知识挖掘与推理规则对已有的各类不同粒度的知识进行关联和语义组织,形成静态知识网络地图;在进行知识服务和实现时可以接受用户需求并形成用户知识需求和知识服务特征,然后建立智能代理程序将接受的用户信息用本体描述语言OWL-S进行本体语义描述、建立高效率的本体语义搜索树实现知识检索、利用推理引擎将知识挖掘与推理规则转化为系统内部的推理过程语言格式以及适当的知识表示与信息推送^[38]。因此,知识服务与实现主要内容包括用户服务接口、知识仓库以及知识服务平台。

5 结 语

综上所述,面向知识服务的知识组织框架是在知识服务目标下和用户需求的驱动下,从宏观视角对知识组织目标、原则、要素、知识组织准备以及知识组织过程进行顶层架构。作者正尝试将构建的知识组织框架应用到水利防汛知识组织中,借助粒度原理描述水利防汛知识的粒度在知识获取与清洗、知识表示与规范、知识挖掘与推理、知识服务与实践等过程中的变化规律。目前作者正在进行从水利防汛需求角度将防汛知识分类和聚类形成知识元、知识单元、文献源等不同粒度大小的知识,初步搭建水利防汛知识组织框架,经过初步应用后虽有个别环节需要完善,但防汛知识组织的总体框架比较合理,下一步在不断收集防汛知识数据基础上,借助本体等技术语义关联各类知识,通过推理挖掘形成知识仓库,最后通过水利防汛知识地图等形式展现给用户,通过为水利防汛部门提供高效的知识服务,最终实现数据知识化、知识有序化和知识服务化目标。

参 考 文 献

- [1] Sue Ellen Wright. Typology for KRRs [C/OL]. NKOS 2008 [2013-02-23] <http://nkos.slis.kent.edu/2008workshop/SueEllenWright.pdf>.
- [2] Zeng M L. Knowledge Organization Systems [J]. Knowledge Organization, 2008, 35 (2-3): 160-168.
- [3] Mai J-E. Actors, Domains, and Constraints in the Design and Construction of Controlled Vocabularies [J]. Knowledge Organization, 2008, 35 (1): 16-30.
- [4] Park On. Opening Ontology Design: A Study of the Implications of Knowledge Organization for Ontology Design [J]. Knowledge Organization, 2008, 35 (4): 209-221.
- [5] June Abbas. Structures for Organizing Knowledge: Exploring Taxonomies, Ontologies, and Other Schema [M]. 100 William St., Suite 2004, New York, NY 10038; (212): 925-8650.
- [6] Pastor-Sanchez J-AMM, Francisco Javier; Rodriguez-Munoz, Jose Vicente. Advantages of Thesaurus Representation Using the Simple Knowledge Organization System (SKOS) Compared with Proposed Alternatives [J]. Information Research: An International Electronic Journal, 2009, 14 (4): 422-432.
- [7] Smith T R, Zeng M L. Building Semantic Tools for Concept-based Learning Spaces: Knowledge Bases of

- Strongly-structured Models for Scientific Concepts in Advanced Digital Libraries [EB/OL]. [2013-6-30]. <http://jodi.tamu.edu/articles/v04/i04/Smith/>.
- [8] Bonome, María G. Analysis of Knowledge Organization Systems as Complex Systems A New Approach to Deal With Changes in the Web[J]. Knowledge Organization. 2012,39(2):104-110.
- [9] Souza Renato Rocha, Tudhope Douglas, Almeida Mauricio Barcellos. Towards a Taxonomy of KOS Dimensions for Classifying Knowledge Organization Systems [J]. Knowledge Organization, 2012;39(4): 179-192.
- [10] Knol[EB/OL]. [2012-11-27] <http://baike.baidu.com/view/1314820.htm>.
- [11] National Center for biotechnology Information[EB/OL]. [2013-5-10]. <http://www.ncbi.nlm.nih.gov/>
- [12] 刘洪波. 知识组织论——关于图书馆内部活动的一种说明[J]. 图书馆, 1991(2):14-18.
- [13] 王知津. 从情报组织到知识组织[J]. 情报学报, 1998(3): 230-234.
- [14] 姜永常. CNKI 数字图书馆知识服务研究[J]. 情报学报, 2004, 23(3): 265-274.
- [15] 王兰成, 敖毅, 李留英. mdKOFT 文献型异构信息源知识组织框架的构建技术[J]. 图书情报工作, 2008, 52(6): 102-105.
- [16] 张文亮, 徐跃权. 论知识组织的三个层次[J]. 图书情报工作, 2011(1): 41-45.
- [17] 贺德方. 国内外知识组织体系的研究进展及应对策略[J]. 情报学报, 2010, 29(6): 963-972.
- [18] 毕强. 数字图书馆知识组织系统建构的发展趋势——从机器可读机器可理解[J]. 国家图书馆学刊, 2010, 19(1): 12-19.
- [19] 薛春香, 朱礼军, 乔晓东. 知识组织系统的描述与评价浅析[J]. 大学图书馆学报, 2010, 3: 75-79.
- [20] 宋培彦, 路青, 赵星. 网络百科知识组织方法研究[J]. 情报资料工作, 2012(5): 73-77.
- [21] 王曰芬, 熊铭辉, 吴鹏. 面向个性化服务的知识组织机制研究[J]. 情报理论与实践, 2008(1): 7-11.
- [22] 王知津, 李培, 李颖, 等. 知识组织理论与方法[M]. 北京: 知识产权出版社, 2009: 25-35.
- [23] 白海燕, 姜波. 基于开源软件构建数字图书馆的知识组织体系[J]. 现代图书情报技术, 2009, 177(4): 7-13.
- [24] 董慧, 姜赢, 王菲, 等. 基于数字图书馆的本体演化和知识管理研究——动态知识组织[J]. 情报学报, 2009, 28(4): 483-491.
- [25] 刘华梅, 侯汉清. 基于受控词表互操作的集成词库构建研究[J]. 中国图书馆学报, 2010(03): 67-72
- [26] 王军, 卜书庆. 网络环境下知识组织规范研究与设计[J]. 中国图书馆学报, 2012, 38(200): 39-45.
- [27] Hobbs J R. Granularity. Proceedings of the Ninth International Joint Conference On Artificial Intelligence. Morgan Kaufmann, 1985: 42-435.
- [28] 卜东波, 白硕, 李国杰. 聚类/分类中的粒度原理[J]. 计算机学报, 2002, 25(8): 810-816.
- [29] 赵昌葆. 工程知识粒度化技术及其应用研究[D]. 西安: 西北工业大学, 2006.
- [30] 杨人子, 严洪森. 基于信息粒度的知识网的模糊分类与检索方法[J]. 自动化学报, 2011, 37(5): 585-595.
- [31] 王知津, 王璇, 马婧. 论知识组织的十大原则[J]. 国家图书馆学刊, 2012, (4): 3-11.
- [32] 常艳. 基于本体的数字图书馆知识组织构建模式研究[D]. 长春: 吉林大学, 2008.
- [33] 马俊, 王光霞, 齐晓飞, 等. 基于场景的不同类别不同粒度知识获取研究[J/OL]. 测绘科学. DOI: CNKI [2013-2-22]. <http://www.cnki.net/kcms/detail/11.4415.P.20130222.1321.004.html>.
- [34] John F. Sowa. 知识表示(英文版)[M]. 北京: 机械工业出版社, 2003: 132-143.
- [35] 温浩, 温有奎. 主题成因的知识元本体转换模型研究[J]. 情报学报, 2011, 30(11): 1123-1128.
- [36] Usama M. Fayyyad, et al. From Data Mining to Knowledge Discovery: Advances in Knowledge Discovery and Data Mining[M]. AAAIIM IT Press, 1996.
- [37] 汪浩祥, 严洪森. 基于多 Agent 可互操作知识化制造动态自适应调度策略[J]. 控制与决策, 2013, 28(2): 161-168.
- [38] 陈红叶, 金国英. 基于本体的茶叶知识服务系统研究与实现[J]. 中国农学通报, 2011, 27(4): 453-458.

(责任编辑 刘志辉)