

本体驱动的知识管理系统模型及其应用研究*

王 昊 谷 俊 苏新宁

摘 要 实施有效的知识管理以提高组织的核心竞争力是知识经济时代下组织重点研究的问题。本体和语义网技术的引入,为知识管理系统的发展和实施带来了新的契机。本文采用模型构建和功能分析方法,对本体驱动的知识管理系统模型进行了总体理论框架设计和 CSSCI 实践应用探讨。研究发现:本体驱动的知识管理系统模型分为三个环节,基于信息处理技术的知识采集,以本体为基础结构的知识组织,以及包括知识地图、知识检索和知识创新在内的以语义网为实现平台的知识应用。以该模型工作原理为指导,设计了一个 CSSCI 学术资源知识管理解决方案,以期 CSSCI 用户的学术研究和科学决策提供有效的知识服务。图 9。参考文献 34。

关键词 知识管理 本体 知识采集 知识组织 知识应用 语义网

分类号 G203

Research on the Model and Its Application of Ontology-driven Knowledge Management System

Wang Hao, Gu Jun & Su Xinning

ABSTRACT Organization should focus on how to implement effective knowledge management so as to improve its core competitiveness under the era of knowledge economy. Introduction of ontology and Semantic Web brought new opportunities for the development and implementation of knowledge management systems. This article, using the methods of model construction and functional analysis, designs the general theoretical framework for the model of ontology-driven knowledge management system and discusses its practical application in CSSCI. The research discoveries that the model of ontology-driven knowledge management system includes three aspects, which is respectively knowledge acquisition based on information processing technology, knowledge organization based ontology as infrastructure, and knowledge applications including knowledge map, knowledge retrieval and knowledge innovation based semantic web as implementation platform. This paper also designs a solution for knowledge management of CSSCI academic resources under the guide of the model working principle, in order to provide effective knowledge services for the academic research and scientific decision-making of CSSCI users. 9 figs. 34 refs.

KEY WORDS Knowledge management. Ontology. Knowledge acquisition. Knowledge organization. Knowledge application. Semantic Web.

1 引言

随着知识经济时代的到来,作为其核心价值的知识凸显出了重要的战略地位。如何有效管理

知识,使其发挥最大的作用,以提高组织的核心竞争力成为了组织重点研究的问题。特别是本体机制和语义网概念的引入和提出,对知识管理的研究产生了重大的影响。由于本体强大的知识表示形式和关联推理机制,基于本体的模型俨然成为了新

* 本文系国家社科基金青年项目“面向语义网本体的知识管理研究”(编号:09CTQ010)的研究成果之一。

通讯作者:王昊,Email: ywhaowang810710@sina.com

一代知识管理系统的自然选择^[1]。本体机制为知识的组织和描述提供了理论基础和技术保障;语义网作为一种新的知识应用模式,为知识 Web 应用的前景描绘出了可操作的蓝图^[2]。

本文试图综合各领域中的知识管理模式,提出一种具有较强操作性的、基于本体、面向语义网的知识管理系统框架,详细描述其组成模块和工作流程,并将其用于指导 CSSCI 学术资源的知识管理研究,以论证该系统模型的可行性、正确性和有效性。

2 本体及语义网在知识管理中的应用

本体用于描述知识的语义,是语义网体系结构中的关键组成部分;语义网本意是使 Web 资源具有机器能够处理的特性,便于用户检索,其实质是使 Web 资源中蕴含的知识显性化。本体的引入和语义网的提出使知识管理的内涵得到了扩展,学者们纷纷将其应用于各自领域的知识管理研究中,基于本体的领域知识管理经历了由模型构建^[3-5]到系统开发^[6-9]的发展过程,提出了数字资源明确性处理、共享性发布、形式化推理的解决方案,并得以在 Web 环境中具体实践。“国共两党关系历史”完整知识库及知识检索系统的开发^[10],更是将我国关于本体驱动知识管理理念的研究推向了高潮,农业、生物学、电信、企业管理^[11]等领域都陆续出现了知识管理模型和系统,且均以本体作为知识组织基础,语义网作为知识检索展示平台,两者逐渐成为知识管理研究的理论基础和技术支柱,甚至出现“离开本体讨论知识管理最终将沦为信息管理研究范畴”的趋势。

在这种背景下,学者们对面向领域的知识管理模型进行概括,总结出了诸多通用模型,描述思路大致可分为三种类型:①从知识流动的角度描述模型的环节组成。这种观点认为基于本体的知识管理研究集中在知识获取、知识存储和知识重用等三个方面^[12],而本体是连接知识流动各环节的核心技术,在知识获取和知识重用中具有重要的应用模式^[13]。在此基础上,有学者从具体实施角度对

知识管理过程进行了细化,提出了一些具有普适性和可操作性的基于本体的知识管理解决方案^[14-15]。②从知识组成角度描述模型的层次结构。基于本体的知识管理系统结构应包括前端应用层、核心集成层和资源层等,用户通过应用层使用集成层中的领域本体和经过标注的文献,这些知识来自资源层的文件系统、数据库服务器及对相关应用的数据封装^[16]。K. W. Chau^[8]将其中的核心集成层细化为描述层和对象层,认为本体在描述层被识别,用户通过应用层实现对对象层中具体知识的使用,实则将本体库分为概念库和实例库两个部分。③从系统开发角度描述模型的功能组件。用户在通讯协议的支持下,通过用户界面请求知识管理应用,基于语义分析和查询、本体推理等中间业务逻辑处理,使用数据储存中的知识资源^[17]。上述功能可被标准化为用户界面、知识管理、知识模型、工作流管理和智能应用系统等组件,本体被存储于知识模型中,结合工作流可实现其在智能系统中的具体应用^[18]。

笔者认为上述研究存在以下几个方面的缺陷:①主要探讨了本体在知识管理流程中的作用,均没有总结出明确的、完整的知识管理系统模型,并对其组成部分进行详细论述;②均提到了语义检索是以本体为基础的,是知识管理的一个功能组成部分,但都没有明确语义检索和语义网之间的关系,即没有明确语义网在知识管理中的地位和作用;③知识本体的来源比较单一,基本上认为形成于知识获取阶段,对于其在知识组织、知识存储、知识重用等其他阶段的丰富和扩充多没有论及。

3 本体驱动的知识管理系统模型构建和分析

通过对各领域知识管理模型及应用的梳理,笔者提出了一个面向语义网的本体驱动的领域知识管理系统模型(见图 1)。整个模型按照知识管理的一般思路^[19]被分成了三个模块,分别是知识采集、知识组织和知识应用。首先对来自不同信息源的信息进行预处理,再采用机器学习或者人工抽取的方式自动或手动地从中采集出具体的知识,包括

实例对象、实例关系以及逻辑规则等,进而将具体知识泛化为抽象知识,如概念知识、概念关系、公理和推理规则等;然后将采集的知识以本体面向对象的形式(如XML等)进行描述,形成领域知识本体库,该知识本体库可用于指导领域信息的语义标注,以获得领域新的实例作为原知识本体的丰富和扩充,本体实现了对领域知识的有效组织,将领域知识连接成为一个巨大的网络知识地图;在此基础

上,对领域本体进行知识应用,以提供高质量的知识服务,具体包括基于知识展示的知识地图服务、基于语义检索的知识检索服务以及基于知识推理的知识创新服务,其中通过知识创新可以获得新的领域知识,作为知识本体的有效扩充,而语义网在该模型中则被认为是知识应用的Web实现方式,是基于本体提供知识服务的有效网络平台。

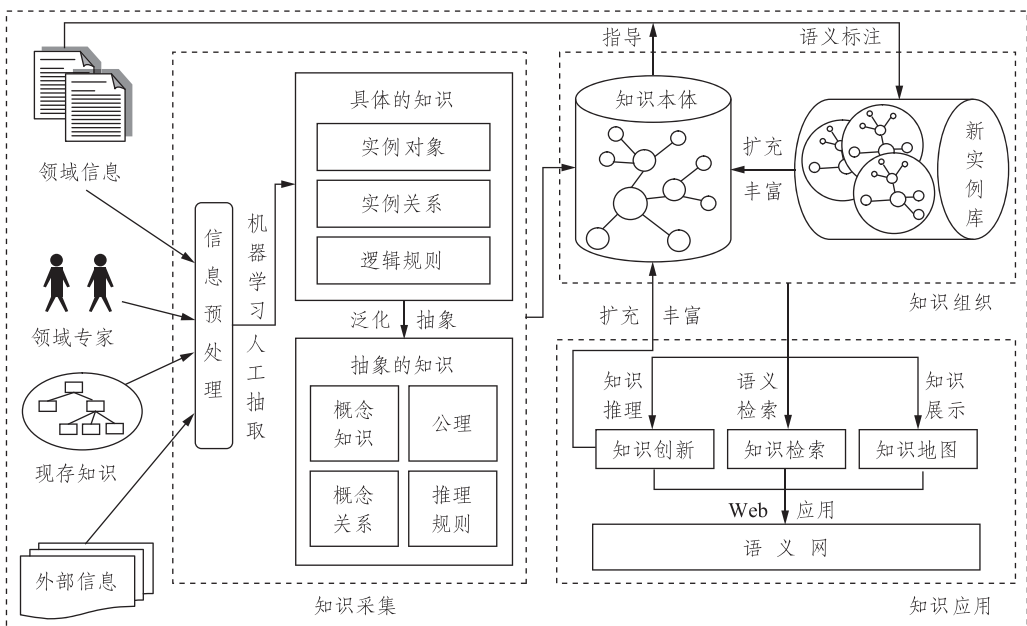


图1 基于本体的领域知识管理系统框架

3.1 知识采集

知识采集模块揭示了知识来源的方式及知识的内涵。领域知识主要来源于四个方面：①领域信息。文献是领域信息存储和传递的主要方式，多表现为文本形式；随着信息资源的数字化发展趋势，领域信息绝大部分被转化为电子信息，很大程度上促进了知识自动采集方式的广泛开展；Internet的迅速发展，领域信息的交流和传递开始趋向于网络方式，其存在形式也逐渐转变为网络结构数据。领域信息是领域知识的主要来源，现多以非结构化文本、结构化数据库、半结构化网页等形式存在。②领域专家。知识是具有思维能力的人对客观存在的思考，最初形成于人脑，后来为了方便交流和传

递才以文献的方式进行记录，有很多知识由于没有及时记录而随着人的消亡而消失。因此，领域专家对现实世界的思考是一种重要的领域知识，是领域知识采集的主要对象之一。③现存知识。随着语言学、知识工程学的发展，研究人员开始进行一些基础性工作，多表现为以人工方式把学科中某一方向的知识进行辨析、明确、梳理、成文，以构建知识库，甚至汇聚大量学者编制通用知识库，因此领域中的现存知识或现存知识中的领域相关内容，包括叙词表、词典、知识库等均可作为领域知识来源。例如以叙词表为基础构建领域本体是目前中文领域本体的主要构建方式^[20-21]。④外部信息。指领域外存在的与领域相关的知识来源，包括时间、地

点、人物、事件等。例如某领域中存在知识“雨花石产于南京”,揭示的是“雨花石”和“南京”之间的关联,那么“雨花石”和“金陵”也有着相同的关系,但是在该领域的信息集合中并不存在这样的关联知识,也没有记录“金陵”和“南京”之间的关系,有时可引入外部信息作为领域知识的有益补充。

不同的信息来源,其文件格式、信息形式甚至存储载体都存在差异,为此需要对不同的信息来源进行预处理,将其转化为统一形式;然后在领域专家的参与下,采用人工抽取或机器学习的方式,从经过预处理的领域内部和外部信息中采集知识。

从总体上说,领域知识可以分为具体的知识和抽象的知识。①抽象的知识,是从众多的领域事实中抽取出共同的、本质的特征所形成的对领域中具体事实的概括性描述,多表现为概念、原理、公式、法则等。例如“作者发表论文”,其中“作者”和“论文”是概念知识,是对具体每一个作者和每一篇文章的抽象,而“发表”则揭示了两者的关系,属于概念关系知识。在本体论中,领域中抽象知识的集合被称为概念模型,包括概念、概念关系、公理和推理规则,由领域中的具体知识泛化获得,可用于指导领域信息的语义标注。②具体的知识,是对领域中现实存在的、在一定时间和地点发生的事实的反映,是抽象知识在现实世界中的具体表现,多表现为对象、事件、规则等。例如“‘王昊’发表‘CSSCI 本体概念模型的构建和描述’”中,“王昊”是具体存在的作者,“CSSCI 本体概念模型的构建和描述”则是“论文”概念在现实世界中的具体表现,两个实例之间具有“发表”关系。在本体论中,领域中具体知识的集合被称为实例库,包括实例、实例关系和逻辑规则,一方面可以通过人工或机器的方式从领域信息中直接采集,另一方面可以在概念模型指导下,通过语义标注形式获取实例知识。具体知识是抽象知识的唯一来源,可以通过机器学习或领域专家,将具体知识自动或手动地转化为抽象知识。

3.2 知识组织

将采集获得的领域知识以一定形式组织成领

域知识库,以备今后进一步使用或实现知识积累,这一过程称为知识组织。知识组织的方式原多采用在信息组织中获得极大成功的方法,例如分类法、主题法、元数据等,叙词表、MARC 等都是其中的典型代表。然而,上述方法虽然能在一定程度上描述词汇间关系,但是知识间语义关联的复杂程度远远超过了它们的描述能力范围。即传统的信息组织方式无法完全描述知识的语义,不能充分表达知识间关联,基于信息组织方式的知识结构需要进一步调整和改善。

随着本体机制在知识工程领域的广泛应用以及面向对象思想的普及,人们逐渐发现以面向对象的方式进行知识组织,以对象属性来描述知识间关系的方法能够充分表达知识语义。例如“作者发表论文”,其中“作者”和“论文”均是领域概念/类,“发表”则是“作者”概念的一个属性,该属性的值域为全体“论文”实例;“被发表”则可以作为“论文”概念的一个属性,该属性的值域为全体“作者”实例。通过这种方式可以充分描述“作者”和“论文”概念的语义关系,并将与实例相关的所有知识聚集在一起。图 2 展示了上述知识的一种具体描述方法。

这种“明确清晰地描述领域概念模型的规范说明,以揭示对象本质”的原理被称为本体,其擅长对领域概念、概念间关系以及概念间约束进行充分的语义描述,逐渐发展成为知识组织方式的典型代表。特别是随着语义网的提出,本体作为语义网中知识描述的核心机制和实现人机语义交流的关键技术,受到了人们的广泛认可,几乎成为知识组织和知识描述的唯一代言。而本体对领域知识揭示机制的具体实现则依赖于一定的形式化语言,XML、OWL 等就是其中的典型代表,图 2 就是采用 XML 的方式来描述领域概念及其关系。本体机制的实质就是使用形式化语言对领域知识进行形式化定义,使得“机器能够理解 and 处理”,从而达到充分揭示知识语义的目的。于是,经过知识采集而获得的领域知识采用本体方式,以 OWL 等形式化语言进行具体描述,进而形成半结构化的知识库。

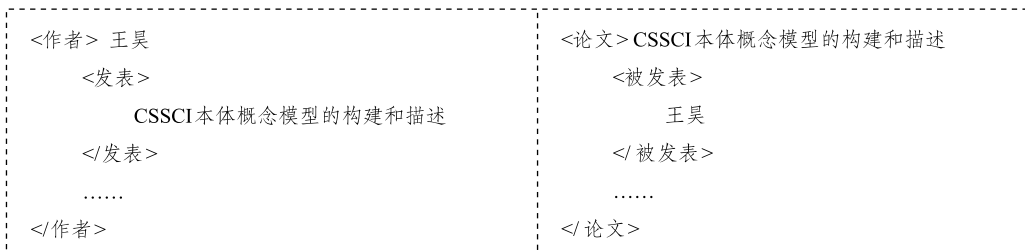


图2 “作者”和“论文”实例知识的一种具体描述方法

经过综合归纳获得的领域知识可以进一步用于指导领域信息的语义标注^[22],语义标注是根据领域本体为领域中的信息及其各组成部分标注所属类别的过程^[23]。语义标注的方法主要有两种:

①人工标注,由领域专家根据本体中的知识定义,借助工具对领域信息的各组成部分添加类别标签。这种方式耗时耗力,仅限于领域资源不是太多的情况。②自动标注,利用自然语言处理技术对领域信息进行词汇语义分析,自动抽取领域词汇,根据关联度计算建立词汇和领域概念、属性和约束类别之间的映射关系。机器自动标注目前被限制在一定应用领域和资源类型范围内,有时还需要辅以一定的人工干预。无论是人工标注还是机器标注,都可以使用一定的语义标注工具加以支持,常用的有SMORE、OntoMat-Annotiser、SHOE、Briefing-Annotizer、GATE等^[24]。语义标注是知识组织模块的重要工作环节,一方面可以获得更多的领域实例知识,丰富并扩充领域知识库;另一方面可以使领域信息获得更准确、更充分的语义解释,提高领域信息的检准率和检全率,加强领域信息被使用的效益和效率。

3.3 知识应用

对领域知识进行组织和描述的最终目的是实现领域知识的共享和应用。笔者对领域知识的应用模式进行了总结,认为主要存在三个应用方向,即知识地图、知识检索和知识创新。

(1) 知识地图

知识地图是由英国情报学家布鲁克斯(B. C. Brooks)提出的一种理论思想,他认为可以将学科

领域中的各知识单元根据其固有联系联结成学科认识地图^[25]。其后,认识地图逐渐演化成为一种以可视化节点方式组织和描述领域内知识单元及其关系的知识管理工具,知识地图的思想、理论、方法 and 应用等都得到了迅速发展^[26-28]。可视化技术的发展和成熟,知识地图逐渐由知识组织理论方法转变为知识展示实践工具,现多聚焦于“以可视化技术展示领域内知识的一种手段”上。因此笔者将知识地图纳入到知识管理体系结构的知识应用环节,用于实现本体中蕴含知识的可视化展示,作为领域知识浏览的可视化导航、知识检索结果的可视化展示以及知识推理过程和结果的可视化描述。

知识地图能够实现知识间导航,提供领域知识浏览的渐进式引导服务。图3(见下页)为某学术资源(仅包含学者与主题)知识地图片段,若开始仅了解学者“苏新宁”的情况,则可以根据该知识地图的引导,发现“科学评价引文分析”为“苏新宁”的研究方向之一,而该方向相关的学者还包括“邱均平”、“周萍”、“杨建林”等,进一步可以导航至“杨建林”的研究主题,发现其中感兴趣的主题“创新力评价”,进而获得该主题的相关学者及其论文,做进一步研究,如此完成了“学者→主题→学者→主题→其他学术资源”的系列导航,实现对所需领域知识的全方位浏览。

从传统意义上讲,知识检索的结果为领域某一知识点,一般可以采用列表方式返回结果,例如检索“苏新宁的研究方向有哪些”,可以列表显示其相关主题,如图4(a)(见下页)所示;然而用户更乐于看到该知识点的上下文相关知识,如这些研究方向的“相关学者”、“相关主题”和“相关论文”等,即

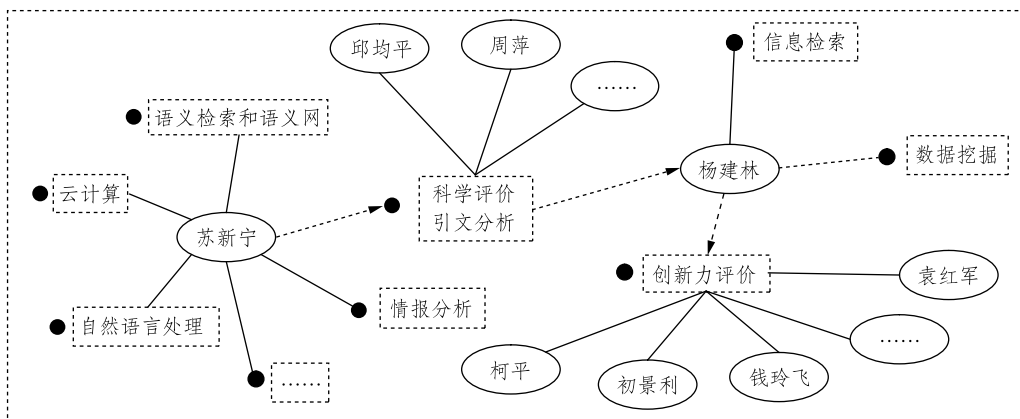


图3 学术资源(学者和主题)知识地图片段

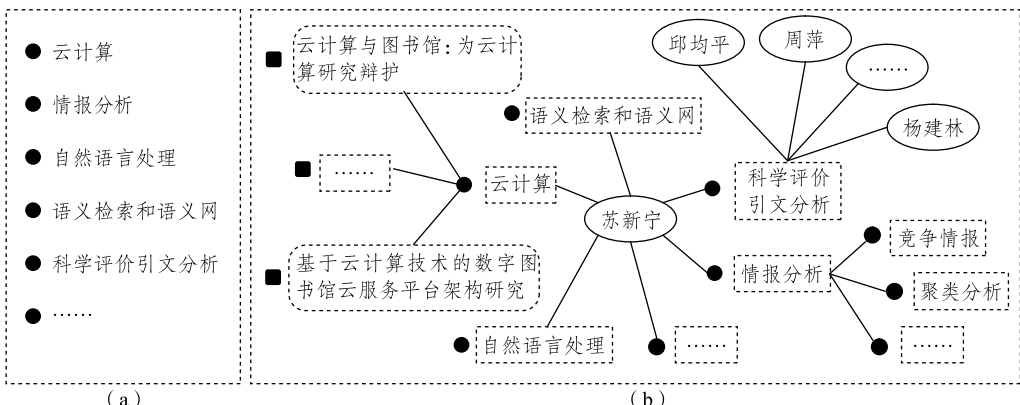


图4 知识检索结果的列表展示和知识地图可视化展示对比

描绘出检索结果的上下文知识地图,如图4(b)所示。以知识地图的可视化方式展示知识检索结果(知识点)及其相关信息(上下文知识)是提高信息服务水平的一种可操作途径。

知识本体期望能囊括领域内所有知识点,然而由于领域中的潜在和隐含知识难以获取,并非直接存储在本体中,而是通过对本体中的知识进行推理和挖掘来发现,知识地图能够实现对知识推理和挖掘过程和结果的可视化描述。如图5(见下页)所示,用户试图了解学者“苏新宁”和“杨建林”之间的关系,然而学者之间的关系知识没有直接存储在本体中,此时可以根据本体规则库中设定的逻辑规则进行蕴含推理。假设存在规则“若a和b的研究主题之间的交叉关联程度达到一定阈值,则a和b

存在主题相似关系(潜合作关系)”,运行上述规则,两位学者之间潜在关系的推理过程及最终结果均显示在图中,其主题相似关系的关联度为5.756,为两者相关主题的交叉关联度之和。

(2) 知识检索

关于知识检索概念,到目前为止还没有统一的定论,不同学者在各自领域都提出了探索性的定义。笔者认为可以从广义和狭义两个层次来探讨知识检索。从广义上讲,知识检索就是综合应用多学科的思想、方法和技术,模拟人类关于知识处理与利用的智能行为和认知方法,实现各种类型信息的准确检索的一种高级信息检索理论与模式^[29];从狭义上讲,知识检索就是将信息资源进行预处理,使之转化为知识资源,并将其按照一定的方式

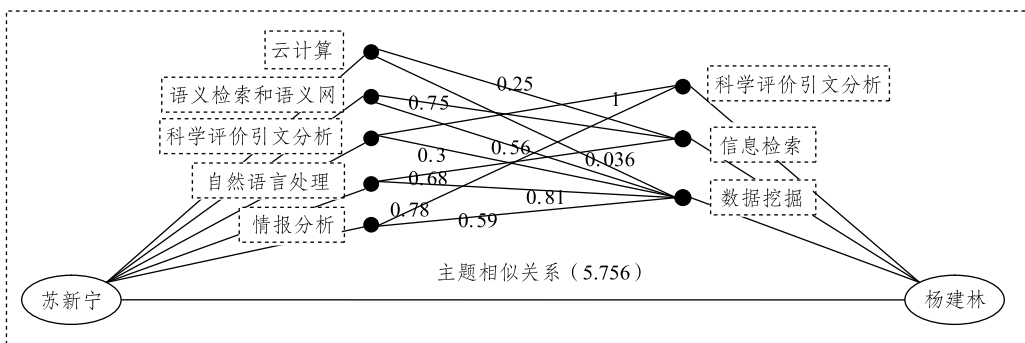


图5 学者关系推理过程和结果的知识地图可视化展示

组织和存储起来,然后按照用户需求利用现代信息检索技术对知识资源进行有效查找的方式和过程^[30]。

本文将知识检索作为知识管理系统模型中的一种知识应用方式,严格上讲属于狭义的范畴,即试图基于知识本体,利用语义检索技术^[31],改善检索服务质量,使用户能够更准确、全面、方便获得所需资源,达到知识检索服务的层次。其内涵包括两层含义:①提供本体中知识资源的检索服务。即采用语义匹配^[32]方式,实现本体库中知识的有效查找^[20]。例如在学术资源本体中,用户试图查找“语义网”主题,如果该主题在知识库中存在,此时不仅返回该主题的释义,更重要的是返回与该主题相关的所有知识点,包括相关的学者、论文、主题、期刊等学术资源,甚至可以列出该主题的上下位和相似主题,以方便用户选择和改变检索策略,提供真正的知识检索服务,使用户能够快速准确地了解和掌握领域知识。②提供基于知识的其他资源检索服务。领域本体构建完成后,还可以将其应用于相同或相关领域的资源检索中。例如将 CSSCI 学术资源本体应用于 CNKI 期刊论文检索系统中,当用户以“语义网”作为检索词时,可以利用 CSSCI 本体获取“语义网”的相关知识,如相似主题“语义检索”、“本体”、相关学者“刘柏嵩”、“罗昊”等作为用户检索策略修正的参考^[33];在获得相应检索结果(期刊论文)后,也可以利用 CSSCI 本体中的相似论文提供检索扩展等。

(3) 知识创新

知识创新是形成领域新知识的所有活动的总称,是一种创造知识的行为。提高领域的知识创新能力、加大知识创新力度是增强领域知识成熟度的重要途径,是实现领域知识管理的重要标志。笔者认为知识创新是基于领域本体进行知识推理的一种知识应用结果,其中知识可作为领域本体的有益补充。例如在学者“合作关系”的描述中,学者之间的合作关系原先并没有存储于领域本体中,基于对学者发表论文的交集运算,可推理发现这一隐含知识;在图5的推理中,若主题间相似度的累加超过了阈值,则认为两者之间存在“主题相似关系”,阈值的人为设定在很大程度上取决于领域专家的经验,而这种经验知识存在于领域本体之外。上述两个例子均实现了一定程度的知识创新,可以进一步将这种通过知识推理获得的学者间的关系纳入到领域本体的范畴中,达到扩充知识本体的目的。

本文对基于本体的知识创新模式进行了归纳:①基于领域内知识推理的知识创新。这种模式实际上就是对领域内潜在及隐藏知识的挖掘和发现。在现有领域知识的基础上,通过综合和比较分析、逻辑和蕴含推理、机器学习和数据挖掘等手段,发现领域内尚未显性化的知识,其知识创新模式如图6(a)(见下页)所示。上述“学者合作关系”的发现就是此类创新模式的典型案例。又如,基于领域中频繁主题之间的相似度,可以构建主题的描述矩阵,进而采用聚类方法得到领域的研究热点,辅以年度参数,甚至可描述出热点的发展轨迹并分析其发展趋势。②基于外部知识影响和作用的知识创

新。其过程如图 6(b) 所示,是在现有领域知识的基础上,通过各种计算方法和推理挖掘技术,在外部知识的影响和作用下,产生新的领域知识的过程。这种外部知识可以来自于另一知识本体,也可以是存在于人头脑中的经验和感觉。“学者主题相似关系”中相似度阈值的设置就是一种外部知识的影响。又如,基于研究单位的相关主题以及主题之间的相似度,可以计算出单位之间的研究相似度,

进而分析单位的研究侧重及合作可能,此时则需要借助外部知识建立研究单位之间的隶属关系,假设“武汉大学信息资源研究中心”挂靠于“武汉大学信息管理学院”,那么前者相关的主题均隶属于后者,此时在借助主题计算“武汉大学信息管理学院”和“南京大学信息管理学院”之间关系时才具有较高的可靠性和准确度。

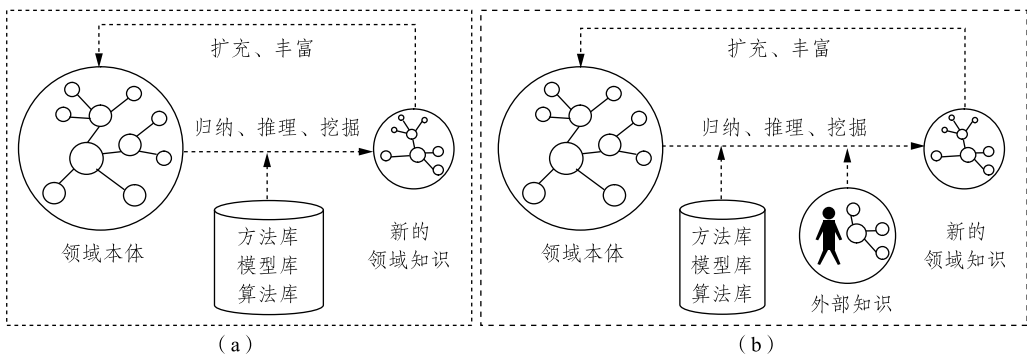


图 6 基本本体的知识创新模式

知识地图是可视化技术在知识服务领域的典型应用,实现了本体知识的顺序浏览和展示;知识检索是检索技术和知识服务的有效结合,不仅能够实现知识的随机浏览,而且借助本体实现了语义匹配,有效提高了检索效率;知识创新则是知识服务的高级模式,借助推理和挖掘技术实现了对隐含知识的发现和新知识的创造。知识地图、知识检索和知识创新一起构成了知识管理系统模型中的知识应用环节。

语义网是新一代的网络平台。Berners-Lee 从 Web 的角度提出了语义网的七层体系结构,认为其不仅是一种网络结构,更重要的是集领域信息采集、知识本体构建、语义检索实现等技术在内的一个完整的 Web 服务环境,是一种能够提高检索效率、改善 Web 服务质量的一种新的 Web 服务模式。本文从知识管理角度出发,认为语义网实质是知识管理和服务网络化的一种工具^[34],是以知识本体为基础,在网络环境中实现知识地图、知识检索、知识创新等的一种具体的知识应用形式。

综上所述,知识管理从体系结构上可以分为

三个环节:基于信息处理技术的知识采集,以本体为基础结构的知识组织,以及包括知识地图、知识检索和知识创新在内的以语义网为实现平台的知识应用。最终形成以知识本体为核心、语义网为操作平台的知识管理系统模型。

4 CSSCI 学术资源的知识管理研究

CSSCI 全称“中国社会科学引文索引”,是一种收录了中国人文社会科学精品期刊相关信息的索引结构。当前 CSSCI 仅从来源文献和被引文献两个角度展示相关学术资源,这对于其他重要学术资源知识的揭示力度不够,无法满足用户快速准确获取知识,了解学术资源之间共现关联的需要。如果能在 CSSCI 数据基础上构建本体,充分揭示学术概念间的相互关系,使之形成知识网络地图,则可以提供更好的知识服务。笔者以知识管理系统模型作为理论指导,对 CSSCI 学术资源的管理流程及其操作方式进行了描述,试图实现 CSSCI 学术资源本体的构建和应用,以期向用户提供更丰富、合理、有

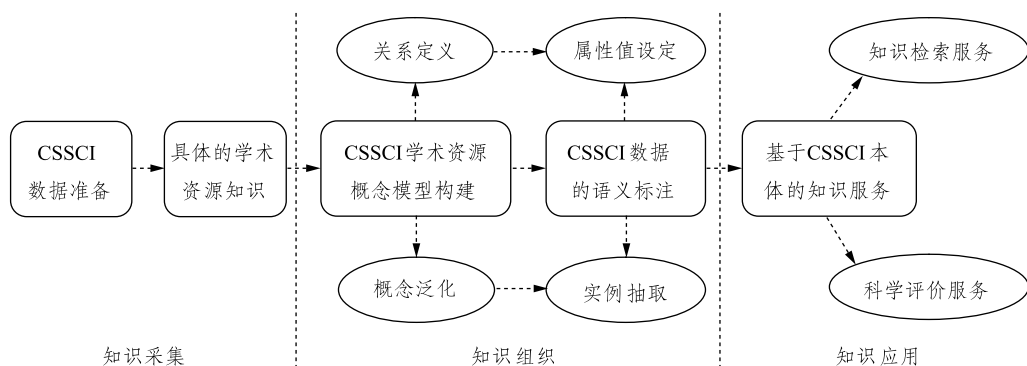


图7 基于本体的 CSSCI 学术资源知识管理流程

效的知识服务,其知识管理系统框架如图7所示。

CSSCI 的知识采集是指从来源数据中抽取出具体的学术资源知识的过程。CSSCI 来源数据以数据库的形式存在,可以根据元数据结构和特点对数据进行分类和综合,基于表字段值、字段值间约束以及元组间约束获得具体的知识,包括学术资源实例及其间关系。例如“苏新宁”、“图书情报档案学”、“情报学报”、“CSSCI 本体概念模型的构建和描述”、“语义网”等均为学术资源的实例知识;“苏新宁 发表 提升图书情报学学科地位的思考——基于 CSSCI 的实证分析”、“中国图书馆学报 隶属 图书情报档案学”、“苏新宁 研究交叉 邱均平”等则均为学术资源实例间关系知识。

CSSCI 知识组织既是对知识采集工作的总结,又是实现知识应用的前提。该阶段的工作主要包括两个方面:①对知识采集的结果进行泛化,从具体的知识中抽象出 CSSCI 本体概念模型,并采用一定的形式化语言对其进行描述。具体的工作包括概念泛化和关系定义,即从学术资源实例中总结出各类概念,然后定义概念的属性作为概念之间的关系。例如“苏新宁”、“邱均平”等均为“学者”概念的实例,“中国图书馆学报”、“情报学报”等均为“期刊”实例等,为“学者”概念定义“发表论文”属性,可以建立“学者”和“论文”之间的关系,为“论文”概念定义“来源期刊”属性,则可以建立“论文”和“期刊”之间的关系。如此,可以建立一个完整的 CSSCI 本体概念模型,在专家评价和实际应用中逐渐趋于完善。②完成 CSSCI 数据的语义标注,具

体工作包括实例抽取、实例属性值的设定、实例的集成以及实例的存储等。这一工作能够实现对具体知识的扩充,使之构建完善的 CSSCI 本体实例库。例如,图8标注了“南京大学信息管理学院”15位学者间的学术关联,是对学者实例间“相关关系”语义解析的一个片段。此外还可由领域专家人工设置一定的推理规则,构建本体规则库实现学术资源知识的蕴含推理。概念模型、实例库以及规则库一起构成了 CSSCI 学术资源本体。

知识应用是体现知识价值的唯一途径,能够为组织的合理竞争和有效发展提供知识服务和决策支持;通过知识应用也能发现知识管理流程中的不足,完善知识采集和知识组织。笔者根据知识应用的三个方向,认为可以从两个方面来提高 CSSCI 知识服务的水平:①基于本体的 CSSCI 知识检索服务。包括三层含义,一是实现对 CSSCI 本体中各学术资源知识点的查询和可视化展示,结合知识地图提供知识定位和知识关联解析服务;二是基于知识间语义关联实现对用户检索表达式和检索结果的知识推荐,提供修正检索策略和查询扩展服务;三是基于推理规则实现学术资源间语义关系的蕴含推理,提供领域知识的扩展和创新服务。②基于 CSSCI 本体的科学评价服务。其主要目的是基于现有知识,结合现代信息分析技术,挖掘领域中的潜在和隐含信息,从而实现学术引导和科学评价。可以从三个角度实现知识再造:一是基于同类学术资源间的两两关系,进行多维尺度和聚类等相关分析,提供学术资源的分类探讨和联合重组服务;二

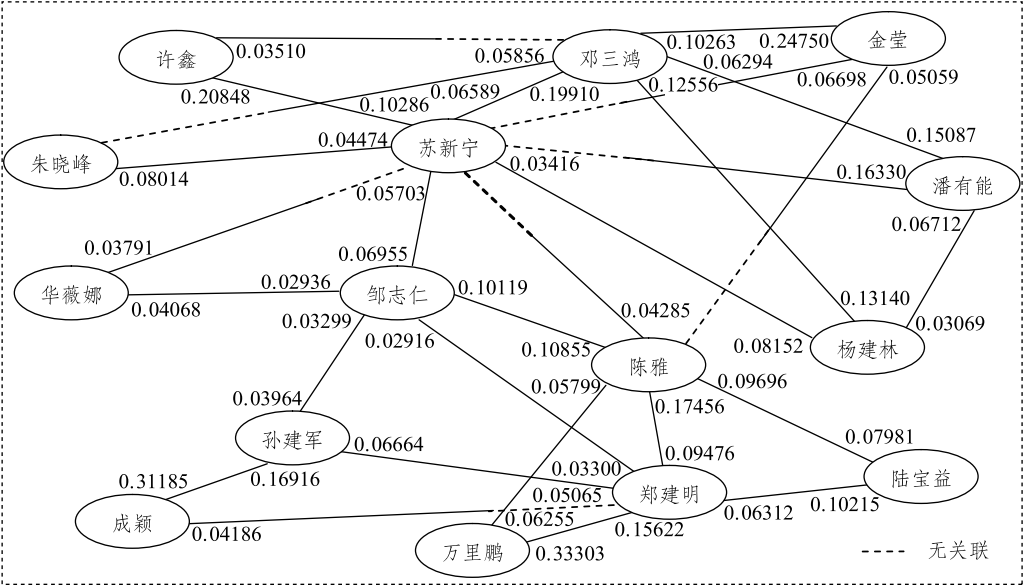


图 8 “南京大学信息管理学院”15 位来源作者间的关联云图

是基于主题间关联以及学者主题和学科主题关系,结合聚类方法,提供学者和学科的研究热点分析和追踪服务,寻找学者间合作甚至学科间交叉的可能;三是在热点探测的基础上,结合时间元素,分析学者和学科研究热点的发展轨迹和趋势,为研

究人员提供研究方向选择、学术焦点确定以及学科发展决策等方向的知识服务。图 9 展示了基于 CSSCI 知识本体(2000—2006)的“图书情报档案学”研究热点总体分布及部分热点的发展趋势状况。

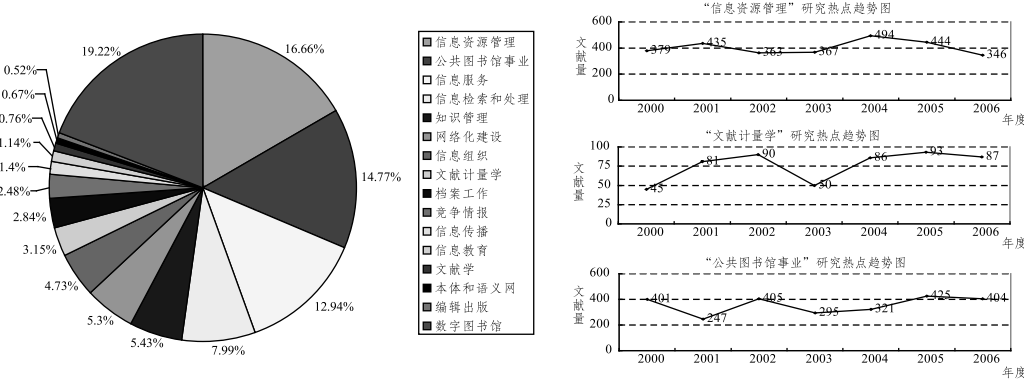


图 9 基于 CSSCI 本体(2000—2006)的“图书情报档案学”研究热点及部分热点发展趋势

5 结语

本文提出了一个比较完善的以知识本体为组

织核心,语义网为应用平台,包括知识采集、知识组织和知识应用等环节的知识管理系统模型。由领域专家手动或机器学习自动方式采集的领域知识,经过本体的明确化、形式化和统一化组织和描述,

以知识地图、知识检索和知识创新等方式实现应用,而语义网则是知识应用的 Web 实现方式。在此基础上,对 CSSCI 学术资源的知识管理过程进行了探讨,提出了一个能够实现学术资源采集、组织和应用的解决方案,以期为 CSSCI 用户的学术研究和科学决策提供高水平和高质量的知识服务。

本文在宏观上阐述了知识管理的一般规律,对学术资源的知识组织结构和知识应用模式进行

了探索;对 CSSCI 本体概念模型的构建、学术资源的语义解析以及 CSSCI 本体的检索和评价服务均没有进行详细论证。随着本体和语义网技术的发展,知识管理具备了实践操作的理论基础和技术保证,CSSCI 学术资源本体化以提供优质的知识服务已经成为用户的迫切需求,上述问题将在今后研究中进一步探讨和解决。

参考文献:

- [1] Razmerita L, Angehm A, Maedche A. Ontology-based user modeling for knowledge management systems[C]//Proceedings of the User Modeling 2003, Johnstown, PA, USA, June 22-26, 2003: 213-217.
- [2] Davies J, Lytras M, Sheth A P. Guest editors' introduction: Semantic-Web-based knowledge management[J]. Internet Computing, 2007, 11(5): 14-16.
- [3] 周义刚. 基于本体的电子政务领域数字档案知识管理系统的设计与实现[J]. 图书情报工作, 2009(15): 129-132. (Zhou Yigang. The design and implementation of digital archives knowledge management system in e-government domain based on ontology[J]. Library and Information Service, 2009(15): 129-132.)
- [4] 俞思伟, 董慧, 姜赢. 基于本体的面向多用户医学知识管理模型研究[J]. 医学信息学杂志, 2009(06): 8-10, 37. (Yu Siwei, Dong Hui, Jiang Ying. Research on multi-user-oriented medical knowledge management model based on ontology[J]. Journal of Medical Intelligence, 2009(06): 8-10, 37.)
- [5] 边文钰, 包振强, 周磊, 等. 基于本体的科研领域知识管理建模研究[J]. 计算机应用研究, 2007(08): 93-96. (Bian Wenyu, Bao Zhenqiang, Zhou Lei, et al. Ontology-based knowledge management modeling of scientific research [J]. Application Research of Computers, 2007(08): 93-96.)
- [6] 夏立新, 徐晨琛, 白华. 基于本体的电子政务知识管理研究[J]. 情报科学, 2009(11): 1607-1611, 1618. (Xia Lixin, Xu Chenchen, Bai Hua. Research on e-government knowledge management based on ontology[J]. Information Science, 2009(11): 1607-1611, 1618.)
- [7] Dieng-Kuntz R, Minier D, Růžicka M, et al. Building and using a medical ontology for knowledge management and co-operative work in a health care network[J]. Computers in Biology and Medicine, 2006, 36(7, 8): 871-892.
- [8] Chau K W. An ontology-based knowledge management system for flow and water quality modeling[J]. Advances in Engineering Software, 2007, 38(3): 172-181.
- [9] Cheng H, Lu Y Ch, Sheu C. An ontology-based business intelligence application in a financial knowledge management system[J]. Expert Systems with Applications, 2009, 36(2): 3614-3622.
- [10] 董慧, 杨宁, 余传民, 等. 基于本体的数字图书馆检索模型研究(I)——体系结构解析[J]. 情报学报, 2006(03): 269-275. (Dong Hui, Yang Ning, Yu Chuanmin, et al. Research on the ontology-based retrieval model of digital library(I)——Explanation of the architecture[J]. Journal of the China Society for Scientific and Technical Information, 2006(03): 269-275.)
- [11] Sureephong P, Chakpitak N, Ouzrout Y, et al. An ontology-based knowledge management system for industry clusters [C]//Proceedings of the International Conference on Industrial Engineering and Engineering Management (IEEM 2007), 2007(12), Singapore.
- [12] 黄新, 徐小娟, 徐国梁. 基于本体的知识管理系统研究[J]. 科学技术与工程, 2005(06): 351-356. (Huang Xin, Xu Xiaojuan, Xu Guoliang. Research of knowledge management system based ontology[J]. Science Technology and En-

gineering, 2005(06): 351-356.)

- [13] 陈韧, 韩永国. 基于本体的知识管理研究[J]. 科协论坛, 2007(01): 1. (Chen Ren, Han yongguo. Research of knowledge mangement based on ontology[J]. Science & Technology Association Forum, 2007(01): 1.)
- [14] 姜瑜, 刘辉. 基于本体的知识管理研究热点分析[J]. 电脑知识与技术, 2009(22): 6086-6087. (Jiang Yu, Liu Hui. Hotspots analysis of research on ontology-based knowledge management[J]. Computer Knowledge and Technology, 2009(22): 6086-6087.)
- [15] 展云. 基于本体的知识管理模式[J]. 无线互联科技, 2010(03): 3-6. (Zhan Yun. Knowledge management models based on ontology[J]. Wireless Internet Technology, 2010(03): 3-6.)
- [16] Benjamins V R, Fensel D, Asuncion Gómez Pérez. Knowledge management through ontologies[C]// Proceedings of the Second International Conference on Practical Aspects of Knowledge Management (PAKM98), 1998(10), Basel, Switzerland.
- [17] Sun Wei, Ma Qinyi, Gao Tianyi. Anontology-based manufacturing design system[J]. Information Technology Journal, 2009, 8(5): 643-656.
- [18] Maedche A, Motik B, Stojanovic L, et al. Ontologies for enterprise knowledge management[J]. Intelligent Systems, 2003, 18(2): 26-33.
- [19] Antezana E, Kuiper M, Mironov V. Biological knowledge management: The emerging role of the semantic Web technologies[J]. Briefings in Bioinformatics, 2009, 10(4): 392-407.
- [20] 董慧, 杨宁, 余传民, 等. 基于本体的数字图书馆检索模型研究(Ⅲ)——历史领域资源本体构建[J]. 情报学报, 2006(05): 564-574. (Dong Hui, Yang Ning, Yu Chuanmin, et al. Research on the ontology-based retrieval model of digital library(Ⅲ)——History domain ontology building[J]. Journal of the China Society for Scientific and Technical Information, 2006(05): 564-574.)
- [21] 颜端武, 岑咏华, 毛平, 等. 领域知识本体的可视化检索研究[J]. 中国图书馆学报, 2007(04): 60-63, 76. (Yan Duanwu, Cen Yonghua, Mao Ping, et al. Visualized searches of domain ontologies[J]. Journal of Library Science in China, 2007(04): 60-63, 76.)
- [22] Uren V, Cimiano P, Iria J, et al. Semantic annotation for knowledge management: Requirements and a survey of the state of the art[J]. Web Semantics: Science, Services and Agents on the World Wide Web, 2006, 4(1): 14-28.
- [23] 张晓林. Semantic Web 与基于语义的网络信息检索[J]. 情报学报, 2002(4): 413-420. (Zhang Xiaolin. Semantic Web and semantic-based networked information retrieval[J]. Journal of the China Society for Scientific and Technical Information, 2002(4): 413-420.)
- [24] 邹亮, 廖述梅. 基于本体的语义标注工具比较与分析[J]. 计算机应用, 2004(s1): 328-330. (Zou Liang, Liao Shumei. Comparison and analysis of semantic annotation tools based on ontology[J]. Journal of Computer Applications, 2004(s1): 328-330.)
- [25] Brookes B C. The foundations of information science. Part IV. Information science: The changing paradigm[J]. Journal of Information Science, 1981, 3(3): 3-12.
- [26] Eppler Martin J. A process-based classification of knowledge maps and application examples[J]. Knowledge and Process Management, 2008, 15(1): 59-71.
- [27] Rao L L, Mansingh G J, Osei-Bryson Kweku-Muata. Building ontology based knowledge maps to assist business process re-engineering[J]. Decision Support Systems, 2012, 52(3): 577-589.
- [28] Semenza Jan C, Höser Ch, Herbst S. Knowledge mapping for climate change and food-and waterborne diseases[J]. Critical Reviews in Environmental Science and Technology, 2012, 42(4): 378-411.
- [29] 张玉峰, 李敏, 晏创业. 论知识检索和信息检索[J]. 中国图书馆学报, 2003(5): 23-26. (Zhang Yufeng, Li Min,

- Yan Chuangye. On knowledge retrieval and information retrieval[J]. Journal of Library Science in China, 2003(5): 23-26.)
- [30] 程慧平, 陈永超. 国内知识检索研究进展[J]. 图书情报工作, 2011(10): 126-129. (Cheng Huiping, Chen Yongchao. Research on development of studying on domestic knowledge retrieval[J]. Library and Information Service, 2011(10): 126-129.)
- [31] Kiryakov A, Popov B, Terziev I, et al. Semantic annotation, indexing, and retrieval[J]. Web Semantics: Science, Services and Agents on the World Wide Web, 2004, 2(1): 49-79.
- [32] Jiang X, Tan Ah-Hwee. Learning and inferencing in user ontology for personalized semantic Web search[J]. Information Sciences, 2009, 179(16): 2794-2808.
- [33] Tran T, Cimiano P, Rudolph S, et al. Ontology-based interpretation of keywords for semantic search[C]// Proceedings of the 6th International Conference on the Semantic Web and 2nd Asian Conference on Asian Semantic Web (ISWC'07/ASWC'07), Springer-Verlag Berlin, Heidelberg, 2007: 523-536.
- [34] Warren P. Knowledge management and the semantic Web: From scenario to technology[J]. Intelligent Systems, 2006, 21(1): 53-59.

王 昊 南京大学信息管理学院副教授, 博士。

通讯地址: 南京市汉口路22号南京大学信息管理学院。邮编: 210093。

谷 俊 上海宝山钢铁股份有限公司。通讯地址: 上海市宝山区富锦路885号。邮编: 201900。

苏新宁 南京大学信息管理学院教授, 博士生导师。

通讯地址: 南京市汉口路22号南京大学信息管理学院。邮编: 210093。

(收稿日期: 2012-08-08)