DOI:10.13530/j.cnki.jlis.190050

基于知识内容的数字图书馆跨学科多粒度知识表示模型构建*

王忠义　夏立新　李玉海

摘　要　在当今大科学时代,为支撑跨学科研究,数字图书馆的知识组织方法需要从基于文献、按学科的知识组织向基于知识内容的跨学科知识组织转变。知识表示作为知识组织的核心与基础,是解决数字图书馆跨学科知识组织的关键。本文首先对馆藏知识资源跨学科的共性与差异特征进行分析,为跨学科的知识表示提供理论支撑;而后,依据知识的抽象程度对馆藏知识进行多粒度划分,识别出知识表示的基本单位;接着,基于对以知识为单元的知识表示方法在跨学科知识表示方面的适用性分析,构建跨学科多粒度知识表示模型,实现馆藏知识资源的跨学科多粒度知识表示;最后,通过案例分析指出该知识表示模型具有多粒度、跨学科、多角度、动态性、语义自然和可扩展等特征,对于支撑数字图书馆跨学科知识组织具有参考价值。图 3。表 5。参考文献 24。

关键词　数字图书馆　知识内容　知识组织　知识表示

分类号　G250. 76

Construction of Interdisciplinary Multi-Granularity Knowledge Representation Model in Digital Library Based on Knowledge Content

WANG Zhongyi, XIA Lixin & LI Yuhai

ABSTRACT

In today's big science era, the demand for interdisciplinary knowledge service is becoming more and more urgent. However, as an important platform for interdisciplinary knowledge service, digital library still organizes knowledge according to disciplines by taking document as unit, which makes it difficult to break the barrier between disciplines, resulting in the dispersion of knowledge about the same thing in different disciplines. In order to support interdisciplinary research, the knowledge organization method of digital library needs to go deep into interdisciplinary knowledge organization based on knowledge content.

As the core and foundation of knowledge organization, knowledge representation is the key to solve the

* 本文系教育部人文社会科学研究青年基金项目“大数据环境下碎片化用户生成内容的多粒度知识组织研究”(编号:19YJC870025)和中央高校基本科研业务费重大培育项目“智慧图书馆管理系统关键技术与应用研究”(编号:CCNU18JCXK04)的研究成果之一。(This article is an outcome of the youth project “Research on Multi-Granularity Knowledge Organization of Fragmented User-Generated Content in the Context of Big Data” (No. 19YJC870025) supported by Humanities and Social Science Research Foundation of Ministry of Education of China and the major cultivation project of the Fundamental Research Funds for the Central Universities “Research on Key Technology and Application of Intelligent Library Management System” (No. CCNU18JCXK04).

通信作者:王忠义,Email: wzywzy13579@163.com, ORCID:0000-0001-8945-783X (Correspondence should be addressed to WANG Zhongyi, Email: wzywzy13579@163.com, ORCID:0000-0001-8945-783X)

interdisciplinary knowledge organization of digital library. To this end, this paper first analyzes the interdisciplinary characteristics of digital library knowledge resources to reveal their commonness and heterogeneity. Secondly, according to the abstraction degree of knowledge, the multi-granularity division of library knowledge resources is carried out, and the basic units of knowledge representation are identified. Thirdly, the applicability of existing knowledge representation methods in interdisciplinary knowledge representation is analyzed. Fourthly, based on the above research, the framework and model of interdisciplinary multi-granularity knowledge representation are constructed to realize the interdisciplinary multi-granularity knowledge representation of digital library knowledge resources. Finally, according to the knowledge representation framework and model, the knowledge of two articles from different disciplines are represented, which not only fully describes the knowledge content of each article but also reveals the interdisciplinary link between them. As a result, the validity of the proposed knowledge representation framework and model is verified. The case study indicates that the interdisciplinary multi-granularity knowledge representation framework and model provided by this paper have the characteristics of multi-granularity, interdisciplinary, multi-angle, dynamic, semantic expression smooth and extensibility, which are of great significance and value to support the interdisciplinary knowledge organization. 3 figs. 5 tabs. 24 refs.

KEY WORDS

Digital library.　Knowledge content.　Knowledge organization.　Knowledge representation.

0　引言

进入 21 世纪以来,科研、教学、人才培养等均呈现学科融合的大趋势,跨学科成为当前知识活动的重要特征。跨学科研究越来越需要图书情报人员帮助解决跨学科知识的存取问题,帮助发现不同学科知识之间的关联,开拓研究视野,这也是当前越来越多的学者们提出的新的知识需求。由此可见,跨学科知识服务是当前人们急需的一种知识服务方式。

在大数据环境下,数字图书馆保存了海量的知识资源,可以成为跨学科知识服务的一个重要平台。然而,当前数字图书馆在实现跨学科知识服务方面仍然存在一些困难与不足,究其原因在于:从馆藏知识资源角度看,资源多学科分布,相同研究对象或科研问题的知识分布在不同的学科之中,相互之间存在学科隔阂(如理论、方法差异,学术语言异构等),形成一个个碎片化的学科知识孤岛,成为跨学科知识组织的一大障碍。从数字图书馆的知识组织方式来看,按学科进行组织,较少考虑知识资源的跨学科特征,并且知识组织粒度过粗,主要以文献为单位进行组织,无法完全揭示不同学科知识内容之间的关联,不足以打破学科之间的隔阂。因此,为支撑跨学科研究,数字图书馆急需在按学科进行知识组织的基础上,深入到知识资源内容层次,对馆藏知识资源进行跨学科的知识组织。而知识表示是知识组织的核心与基础,恰当的知识表示方法是解决数字图书馆跨学科知识组织的关键,直接决定了上层的跨学科知识处理的效率、跨学科知识组织的方式和跨学科知识服务的方式。由此可见,数字图书馆若要深入到知识资源内容层次,实现跨学科知识组织,就需要探索更加深入的知识表示方法。然而,当前的知识表示方法尚不能完全支撑知识内容层次上的跨学科知识表示,为此,本文以数字图书馆馆藏知识资源为研究对象,探索知识内容层次上的跨学科知识表示问题,以期为数字图书馆跨学科的知识组织和精准知识服务

提供支撑。

1 相关研究

布利斯(H.E. Bliss)于 1929 年首次提出知识组织的概念[1],随后知识组织便得到快速发展。知识表示作为知识组织的基础,伴随着知识组织的发展也不断深入。从纵向上看,国内外知识表示研究经历了三个发展阶段。

(1)以文献为基本单元的知识表示

由于受历史和科学技术的影响,知识表示在早期主要关注文献的类别表示,知识表示的方法主要有分类号和主题词等。

基于分类号的知识表示方法以分类号作为知识标识,以等级结构显示文献主题概念间的关系,按学科体系排列类目。分类号这种根据学科性质对文献进行系统化表示的方式,使得有关同一研究对象的知识由于隶属于不同学科而被存储在不同学科类目下,因此,人们很难借助分类号全面检索到跨学科的有关某一研究对象的所有文献,成为人们在不熟悉的学科中检索到相关文献的一大障碍[2-3]。

基于主题词的知识表示是以表达文献主题内容的词语作为标引对象,对文献中的知识内容加以揭示的方式。具体包括:标题词、元词、叙词等。基于主题词的知识表示方法可以聚合主题相关文献,具备跨学科知识表示能力。然而由于不同学科的术语异构问题,难以借此发现所有跨学科的知识资源;此外,主题词是受控词,也不适合对跨学科的知识资源进行自动标引[4]。

(2)以信息为基本单元的知识表示

信息单元是指以文献的某些外部和内容特征信息作为知识表示的基本单元。它不是文献本身,只是用来控制和处理文献的信息标识,包括元数据法、关键词法等。

元数据是关于数据的数据,是专门用来描述文献特征的数据。元数据作为一种描述知识资源基本特征的知识表示方式,可被用来进行知识标引[5-6]。然而基于元数据的知识表示方式不能完全解决语义异构问题,更不能系统揭示知识之间的语义关系,因此元数据难以揭示不同学科知识之间的内在关系。

基于关键词的知识表示方法是指以关键词来描述知识主题的知识表示方式。关键词属于自然语言,普遍存在同义词、近义词现象,在识别使用不同术语进行标引的跨学科知识时存在不足,而且由于缺乏对知识资源的语义揭示,如果不知使用什么术语去检索存在于其他学科中的知识,人们将难以发现跨学科的知识资源。

(3)以知识为基本单元的知识表示

该方法以知识单元为知识表示单位,不仅注重揭示文献单元的学科、主题特征,更注重揭示文献所载荷的知识内容本身。具体包括知识地图、谓词逻辑、语义网络、框架、知识本体、语义网等。

知识地图类知识表示方法包括知识地图[7]、主题地图[8]、概念地图[9]等。知识地图类知识表示方式兼具知识内容与知识载体揭示双重功能,具备跨学科知识表示能力。但知识地图类的知识表示方法主要被作为知识资源目录,在揭示知识资源内容语义方面的功能未得到充分发挥,因此在跨学科知识组织方面存在一定的局限性。

基于逻辑的知识表示包括命题逻辑、一阶谓词逻辑和高阶逻辑。基于逻辑的知识表示是陈述性的,可以实现知识表示和知识推理的分离,从而使得知识推理不依赖于特定的学科领域,因此具有一定的跨学科知识表示功能。然而,这种知识表示方法难以表示过程性知识,在对不确定性知识进行表达时也存在一定的局限性。

语义网络[10]将知识表示为节点与边构成的模型,是一种用有向图表示知识的系统。基于语义网络的知识表示方式使用统一的三元组形式对知识内容加以描述,能够做到知识表示的领域无关,因此可以为跨学科知识表示提供借鉴。然而语义网络对节点和边的描述没有统一

的标准，从而不便于知识分享，此外语义网络不适用于表达过程性、控制性的动态知识，因此在对包含大量动态知识的馆藏资源进行跨学科知识表示时具有局限性。

基于框架的知识表示方法[11]以一种类似框架的结构来对知识进行建模，表示事物的属性以及事物之间的类属关系。脚本作为另一种形式的框架知识表示方法[12]，主要用来表达动态的过程型知识。框架可以用于对不同学科中的通用知识的表达，然而在表达动态的过程型知识方面能力不足；脚本虽然可以对动态的过程型知识进行表示，然而不具备对事物基本属性的描述能力，也难以描述事件发展的所有可能方向。因此基于框架的知识表示方法虽然可以为跨学科知识表示提供支撑，但仍需进一步优化。

知识本体是指对概念体系明确的、形式化的、可共享的规范说明[13]。由于知识本体能够为不同学科之间的高层互操作提供良好的基础，因此可以解决跨学科的术语异构问题。然而知识本体强调的是抽象概念表示，不关注具体的实例信息。此外知识本体还忽略了概念知识与实例知识之间的关联。可见知识本体在进行跨学科知识表示时存在一定局限性。

语义网[14]发展至今先后出现两种最佳实践：关联数据[15]和知识图谱。关联数据不仅可为知识表示提供语义支持，而且能够建立知识单元的任意关联[16]，细化知识表示粒度。知识图谱作为关联数据的进一步发展[17]，以符号形式描述现实世界中的概念、实体及其相互关系。由于关联数据和知识图谱均采用 RDF 描述知识资源，且具有通用性，因此可以支撑数字图书馆跨学科的知识表示。此外它们均具有对知识资源细粒度的表达特性，可以实现对馆藏资源的细粒度表示。然而关联数据本身不会给知识增加任何语义信息，只是携带语义数据，而且关联数据无法解决语义异构问题。知识图谱目前也只能表示以实体为核心的事实型知识，无法对数字图书馆中的常识知识、情境知识等进行表示，并且知识图谱描述的知识是静态的，无法描述动态的知识内容。

综上，当前数字图书馆知识表示方式正处在以信息为基本单元的知识表示方式为主，以文献为基本单元的知识表示方式为辅的阶段，并朝着以知识为基本单元的知识表示方式迈进。由于信息单元的处理最终还是必须依赖文献单元，因此仍然属于粗粒度的知识表示方式。虽然人们也提出了一些以知识为基本单元的知识表示方式，但发展还不成熟。而且，在文献层次进行知识表示很难打破不同学科之间的隔阂。数字图书馆若要实现跨学科的知识表示方式，急需将知识的控制单位由物理层次的文献单元深入到认知层次的知识内容单元。而知识内容单元具有多粒度性，这就意味着跨学科的知识表示势必要建立在多粒度知识单元的基础之上。基于上述分析，本文提出数字图书馆跨学科多粒度知识表示模型，以期为数字图书馆跨学科的知识组织提供支撑。

2　数字图书馆馆藏知识资源的跨学科特征分析

在对数字图书馆的馆藏知识资源进行跨学科知识表示时，若要真正打破学科主题的界限，就需要对文献中的具体知识内容进行描述，揭示所有学科领域中知识内容的共同特点和不同学科领域中知识内容的特殊性。为此，本文首先对馆藏文献中知识内容的跨学科特征进行分析，明确其共性和异性特征，为跨学科知识表示提供理论支撑。

学科知识之间的共性在于：一方面，从认知对象来看，同一事物的本质是确定的和统一的，人们虽然从不同的学科角度对该事物进行研究，形成了描绘该事物的一个个“平面图景”，但真实的世界是立体的，因此，不同学科所获得的关于同一事物的知识必然存在较强的相关性，它们共同绘制了一幅立体的世界图景，即千差万别的学科知识统一于研究对象（即物质），具有相关性。另一方面，从认知过程来看，人类对

自然界、人类社会以及人类心理世界认知的最终目的，是为了实现对事物的整体把握和深刻理解，因此，无论是从横向还是纵向来看，不同学科在研究范式上具有一定的通约性。学科知识之间的差异在于：同一研究对象，不同学科关注的科研问题往往不完全相同，进而导致解决问题时所采用的研究视角、研究方法、理论体系、学术语言以及得到的研究结论等也不尽相同，因此，同一对象的相关知识在不同学科中的表现千差万别，各有其价值和地位，即学科知识之间存在广泛的差异性。由此可见，不同学科领域知识的共性在于它们的研究对象相同，研究范式或内容具有相关性；差异在于它们关注的具体科研问题不完全相同，解决问题时所采用的理论、研究方法、学术语言以及研究结论等也不尽相同。如对于同一研究对象“人”，医学研究人的身体健康问题，心理学研究人的心理健康问题，虽然研究对象相同，研究的科研问题在总体上相关，即都是“人的健康问题”，但具体科研问题不同，一个是“身体健康问题”，一个是“心理健康问题”，由于研究的具体问题不同，因此，采用的研究方法、理论、术语和得出的结论也肯定存在一定的差异。

基于上述分析，为打破学科知识之间的隔阂，本文提出在对数字图书馆馆藏知识资源进行多粒度知识表示时，应该以文献中的知识内容为揭示对象，以文献中的研究对象为中心，根据文献中研究的具体问题对不同学科的知识加以区分，并基于文献中各种粒度的知识单元之间的语义相关性，揭示不同学科知识之间的关联，只有这样才能将来自不同学科的相同研究对象的知识关联起来，实现馆藏知识资源的跨学科知识表示。为此，本文从研究对象、科研问题、知识单元三个层次，对数字图书馆馆藏知识资源进行跨学科的多粒度知识表示。

3 数字图书馆馆藏知识资源中知识单元的多粒度划分

知识表示建立在知识分类的基础之上。对馆藏知识资源进行分类有利于更为深入、更为明确地对馆藏知识资源的特征进行描述和揭示。要打破不同学科之间的隔阂，实现跨学科知识表示，需要将知识的控制单位深入到文献中的知识内容单元，而从文献单元到文献中最细粒度的知识单元之间存在多种不同粒度的知识单元，为此，本文基于对馆藏知识资源跨学科特征的分析，依据知识的抽象程度对馆藏知识资源进行多粒度的划分，提出多粒度知识分类方式（见图 1），将其划分为实例知识、概念知识、事件知识和事理知识四个粒度层次，以识别出数字图书馆跨学科多粒度知识表示的基本知识单元。

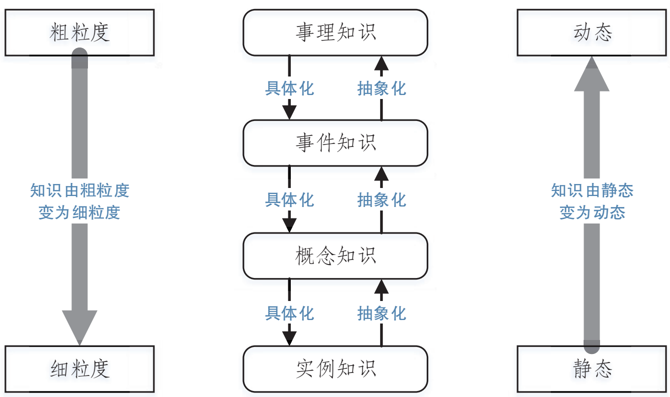

图 1 数字图书馆馆藏知识资源的多粒度知识分类

(1)实例知识

实例是世界上存在的一切具体事物。每个实例都具有特定类型,这种类型由抽象的概念来描述。例如"人"是一个概念,"屠呦呦"是"人"这一概念的一个具体的实例。基于上述分析,本文将实例知识定义为:由实例对象以及实例对象之间的关系构成的知识内容,是最细粒度的知识单元。实例知识是在现实世界中出现的示例,是概念知识的具体体现。实例知识具体包括实例对象、实例对象属性、实例对象之间的关系等内容。

(2)概念知识

从逻辑学角度来看,概念是人类思维的基本单位,是对实例对象特有的本质属性的抽象揭示和描述,是对实例对象内涵的把握,具有抽象性和概括性。这里的实例对象既包括具体的实例对象、实例对象的集合,也包括实例对象的各种性质和关系等。概念由内涵和外延两部分构成[18]。概念的内涵是指概念的含义,由概念所反映的客观对象的特有属性构成,即概念的客观内容。概念的外延是指概念适用的对象或范围,由概念所指称的客观对象或范围构成。

基于上述分析,本文将概念知识定义为:人们对实例知识进行归纳、概括所形成的,用于揭示实例对象本质内涵的知识内容。概念知识可以帮助人们认识实例的同一性和与其他实例的差异性。没有概念知识就无法进行判断和推理。概念知识由概念标识(即概念名称)、概念内涵(即属性)和概念外延(即实例)构成。概念标识通过语词来表达,语词是概念的物质承担者,概念则是语词的内涵,两者是形式与内容的关系;概念知识的内涵是对概念质(是什么)的规定;概念知识的外延是对概念量(有什么)的规定。

(3)事件知识

事件起源于认知科学,随后在哲学、语言学、计算机科学等领域得到广泛的研究和探讨。但当前尚未形成关于事件的统一定义。不同的领域对事件有着不同的理解和界定。认知科学认为人类是通过对事件以及事件之间关系的观察来认识世界的,因此,将事件定义为世界上发生的所有事情[19]。从哲学视角来看,事件是现实世界中事物运动的反映,是发生在特定时间、情境下的,由若干角色参与的,具有特定动作特征的事情[20]。语言学家将事件理解为在特定时间、地点发生的包含一个或者多个参与者的事情[21]。从计算机科学的角度来看,事件是动词及其涉及的相关语义元素所描述的语义信息[22]。由此可见,事件是一个泛指的概念,不仅包括人类社会中发生的事情,还包括自然界、人类精神世界等发生的一切事情。

从上述定义可以看出,尽管各学科领域从不同视角对事件进行了界定,且有着不同的理解,但它们在事件的构成要素上基本达成共识,都认为事件元素包含时间、空间、事件参与者、参与者的动作行为以及伴随的状态改变等要素。综合上述分析,本文将事件知识定义为:由描述事件的时间、空间、参与者、行为和状态等概念以及概念之间的语义关系构成的知识内容。事件知识通常包括:事件知识标识、事件知识元素、事件知识角色、事件知识来源构成。事件通常由动词触发,因此事件知识主要使用动词加以标识。事件知识元素包括事件发生的具体时间、空间、事件的参与者、参与者的行为等。事件知识角色揭示的是事件知识元素与事件之间的语义关系,指出事件元素在事件中扮演的角色。事件知识来源是指馆藏知识资源中对事件的自然语言形式的描述,同一事件知识可以有不同的事件知识来源。人们在科学研究的过程中,为了解决某些科研问题,通常会在特定理论的指导下,采取一定的方法来解决这些问题。在该过程中常常伴随着各种类型的事件发生,从而形成事件知识。例如在竞争情报工作中,就涉及竞争情报采集事件、竞争情报处理事件、竞争情报组织事件、竞争情报分析事件、竞争情报服务事件等类型的事件知识内容。

(4)事理知识

"理"指"规律"。事理,也即事情背后蕴含

的道理,是对事件之间规律性逻辑关系的揭示和描述,这些关系包括因果、条件、反转、并发、组成、上下位等。世界上一切事物都存在普遍联系,遵循一定的规律发展变化。事理是人们在实践过程中分析、归纳出来的事物发展变化的规律、道理,是连接实践对象和实践目的的桥梁,以指导人们把"事"做好,增强人们做事的能力。

基于上述分析,本文将事理知识定义为:用来揭示特定背景下自然界、人类社会、人类精神世界中发生的事件背后蕴含的道理、规律以及由此产生的原则、方法等共同构成的知识内容。事理知识是一种用来解释"为什么"(Why)和"如何做"(How)的知识,用于描述问题的求解过程。依据适应范围,事理知识又可以划分为具体事理知识和一般事理知识两个层次。具体事理知识是指从某一类活动中提炼出来的规律和道理,与活动的主体和客体密切相关,具有较强的情境依赖性;一般事理知识是指从人类所有活动中抽象出来的适用于"一切事"的规律和道理,与具体的主体和客体无关,情境依赖性较低。事理知识主要由事理知识标识、事件、事件逻辑构成。事理知识标识通常由事理知识所指向的具体科研问题来标识;事件即事理知识中涉及的事件知识;事件逻辑即在运行规律、原则指导下,事件在时间上的先后和衔接关系,这种先后和衔接关系将一系列相关事件组织成有机关联的事件链。通过对数字图书馆馆藏知识资源的特征分析发现,一篇文献通常会论述一个核心科研问题,而为便于分析和解决该核心科研问题,往往又会划分为多个子科研问题,因此,同一文献中可能存在一个围绕核心科研问题的最长事件链(主事件链)和围绕子科研问题的多个短事件链(辅助事件链),这就决定了文献中的事件链应该是一个具有层次结构的复杂事件链。

在四种粒度的知识中,实例知识、概念知识属于静态陈述型知识范畴,能够回答"When、Who、What、Where"等问题;而事件知识、事理知识属于描述事物逻辑的动态程序型知识范畴,可以回答"Why、How"等问题。静态陈述型知识需要动态程序型知识依据逻辑加以运用,反过来,静态陈述型知识又可以为动态程序型知识提供必要的基本事实资料,两者相辅相成,共同构成了整个知识体系。

4 数字图书馆馆藏知识资源的跨学科多粒度知识表示模型构建

本研究首先对现有的以知识为基本单元的知识表示方法进行分析,而后基于知识单元的多粒度划分,提出数字图书馆跨学科多粒度知识表示框架,构建数字图书馆跨学科多粒度知识表示模型。

4.1 以知识为基本单元的知识表示方法比较

依据知识表示所采用的编码方式的不同,以知识为基本单元的知识表示方法可以划分为符号主义和联结主义两种类别。符号主义的知识表示是通过图形、概念、命题等使知识符号化、结构化的过程;联结主义的知识表示是通过类比人脑神经活动过程而产生的,通过定义大量神经元节点构建复杂网络对知识进行表示的过程。符号主义的知识表示方法以离散的物理符号表示概念、命题等知识的结晶,可理解性强;联结主义的知识表示方法不对知识进行明确的编码,而是以数学模拟的方式将知识表示为特定时刻的网络活动模式,它所表示的是尚未结晶的知识,可解释性较差。数字图书馆中的馆藏知识资源是人们用文字表达出来的知识结晶,因此,符号主义的知识表示方法比较适用。为此,本文对符号主义的知识表示方法从优点、缺点、可表示的知识类型以及支持多粒度知识表示的程度等方面进行比较(见表1),可为数字图书馆馆藏知识资源的跨学科多粒度知识表示提供方法上的支撑。

表 1　以知识为基本单元的知识表示方法比较

表示方法	优点	缺点	可表示的知识类型	支持多粒度
谓词逻辑	简单、自然、精确、灵活、易于实现。	推理过程冗余、效率低,易出现组合爆炸。	陈述型知识 程序型知识	一般
框架	具有良好的继承性和一致性。	不善于表示程序性。	陈述型知识 程序型知识	很好
脚本	可以描述固定的事件序列。	不具备元素基本属性的描述能力。	程序型知识	一般
语义网络	简单,知识表示的结构统一。	知识表示不严格,无法消除二义性,知识处理复杂。	陈述型知识	一般
语义网/本体	知识表示的结构统一,便于知识共享。	难以实现。	陈述型知识 程序型知识	很好
知识图谱	知识表示的结构统一,便于知识共享。	不太关注对知识框架的定义。	陈述型知识	一般

由表 1 可以看出,各种知识表示方法均具有各自的优点和缺点以及适用的知识类型,并且在支持多粒度知识表达的能力上也存在差异。需要指出的是,无论是脚本、语义网络、本体、语义网还是知识图谱,它们在本质上都可以看作是基于槽的框架法,其区别在于槽是否具有层次、时序、控制等关系。其中,语义网络、本体、语义网和知识图谱是没有层次关系的、更加灵活的、基于槽的框架法,在拓扑结构上可以看成是一个图或网络。脚本是具有时序和控制关系的框架,由此可见,它们与框架法没有本质区别,只是框架法的一般化和特殊化。基于上述分析,本文根据各种粒度的知识单元的特征,提出基于框架的数字图书馆馆藏知识资源跨学科多粒度的知识表示方法。具体来说,基于知识图谱对实例知识进行表示,基于框架法对概念知识、事件知识进行表示,基于脚本对事理知识进行表示,综合各种知识表示方法的优点,对各类知识表示方法的缺点进行改进和优化,从而提供更加细致、复杂的知识描述,最终形成一个框架系统,该框架系统等价于一种复杂的语义网。

4.2　跨学科多粒度知识表示框架

在上述研究的基础上,本研究首先搭建数字图书馆馆藏知识资源的跨学科多粒度知识表示框架(见图 2),该知识表示框架能够忠实于不同学科知识本身的特征。首先,该知识表示框架以研究对象为中心,以具体科研问题为学科区分,可以有效揭示跨学科知识的共性和差异特征;其次,该知识表示框架通过知识粒化,对知识进行多粒度表达,即事理知识、事件知识、概念知识和实例知识,通过多粒度的知识表示,既可以有效描述知识的多粒度关联特征,也保证了对知识的完整表达;接着,该知识表示框架引入情境(时间、地点),不仅有助于明确主体和客体的语义和行为,更重要的是可以揭示知识在不同学科之间的动态演化特征。

4.3　跨学科多粒度知识表示模型

跨学科多粒度知识表示框架揭示了文献中包含的各种粒度大小的知识单元之间的粒化和关联关系,为各粒度大小的知识单元的知识表示提供了依据。本文通过对共性知识的分析发现,不同粒度大小的知识单元大都具有三个基本属性:内部属性,是对知识单元内部包含的知

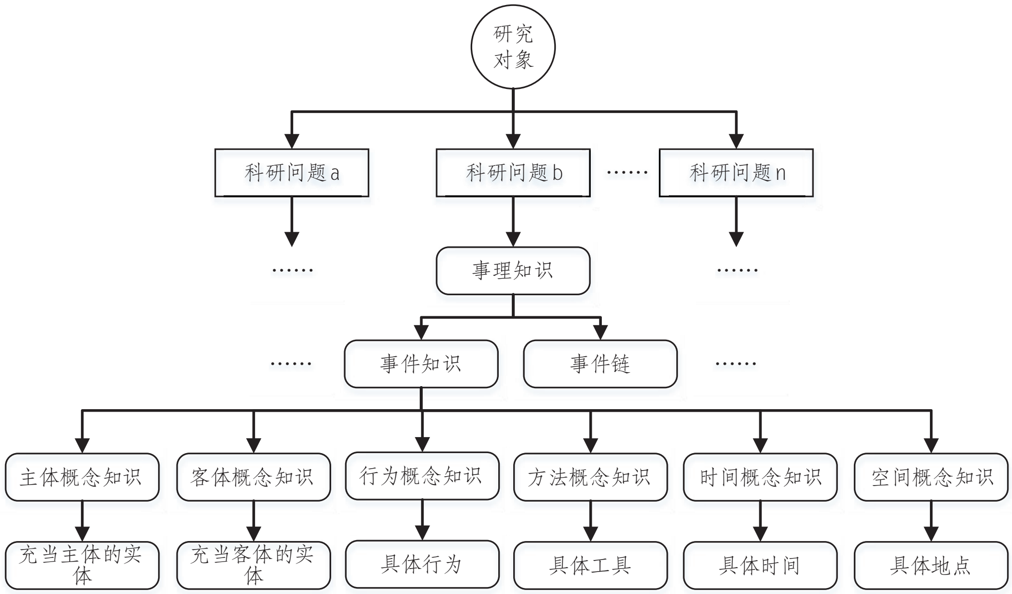

图2　数字图书馆馆藏知识资源的跨学科多粒度知识表示框架

识元素及其内部关联的描述;外部关联属性,揭示知识单元之间的相互作用,是对知识单元之间关系的描述;情景属性,反映一个知识单元所处的情境信息,是对外部情境关联的描述。基于对知识单元属性的分析,本文构建了各种粒度大小的知识单元的表示模型(分别见表2、表3、表4、表5)。

由表2可知,事理知识表示模型主要对其涉及的研究对象、科研问题、事件、事件链、时间、空间、关联和来源进行了描述。研究对象是指科学研究所指向的实体事物;科研问题是指科学研究所指向的实体对象的某个方面;事件是指事理知识中所涉及的事件;事件链,即事件的流程,由事件之间的各种逻辑关系构成;时间和地点描述了事理知识涉及的相关时间和地点信息,事理知识的发布时间和地点有助于揭示事理知识的演化发展过程;关联揭示了不同事理知识之间的外部联系;来源指出了事理知识的文本描述或文献出处。事理知识主题标识由科研问题的名称构成,用来揭示事理知识的主题的内容。

由表3可知,事件知识表示模型定义了事件

表2　事理知识表示模型

主题标识	名称::={事理名称}
内部属性	研究对象::={研究所指向的事物} 科研问题::={研究所指向的方面} 事件::={包含的事件} 事件链::={事件流程}
情境属性	空间::={描述事理知识涉及的地点} 时间::={描述事理知识涉及的时间}
外部关联属性	关联::={与其他事理知识之间的关系} 来源::={描述事理知识的文本及文献来源}

表3　事件知识表示模型

主题标识	名称::={事件名称}
内部属性	主体::={充当事件主体的概念} 客体::={充当事件客体的概念} 方法::={描述事件凭借手段的概念} 行为::={描述事件行为的概念} 事理::={事件所属的事理}
情境属性	空间::={描述事件发生地点的概念} 时间::={描述事件发生时间的概念}
外部关联属性	关联::={与其他事件知识之间的关系} 来源::={描述事件知识的文本及文献来源}

表4　概念知识表示模型

主题标识	名称::={概念名称}
内部属性	内涵::={描述概念本质内涵的属性} 外延::={与概念相对应的实例对象} 事件::={概念所属的事件及在事件中所扮演的角色}
情境属性	空间::={描述概念知识涉及的地点} 时间::={描述概念知识涉及的时间}
外部关联属性	关联::={与其他概念之间的关系} 来源::={描述概念知识的文本及文献来源}

表5　实例知识表示模型

主题标识	名称::={实例名称}
内部属性	属性值::={实例对象特定属性的具体属性值} 概念::={实例所属的概念}
情境属性	空间::={实例对象所处的具体地点} 时间::={实例对象存在的具体时间}
外部关联属性	关联::={与其他实例之间的关系} 来源::={描述实例知识的文本及文献来源}

知识包含的事件元素和外部关联。事件元素在事件中扮演的角色定义为六种类型:主体、客体、行为、方法、时间和空间。主体是事件的发起者,又可细分为施事、感事、致事、主事等类型;客体是事件的承受者,又可细分为受事、与事、结果、对象、系事等类型;行为是事件涉及的动作,由事件的触发词构成;方法在这里是一个宽泛的概念,泛指借助的一切事物,具体包括工具、材料、方式、技术等类型;时间是指事件发生的时间,又可细分为开始、结束、间隔等类型;空间是指事件发生的地点,具体又可细分为起点、终点、范围、场所等类型。事件知识的主题标识由事件触发词来表示,是对事件类型的标识。事件知识的外部关联属性主要用于描述事件知识之间的联系,并指出事件知识的文本或文献来源。

由表4可知,概念知识表示模型主要从概念的内涵、外延、事件、时间、空间、关联和来源几个角度来对概念知识进行描述。内涵是对概念本质内容的揭示,由属性构成,是对概念知识质的规定;外延是对概念内涵所适用的对象范围的揭示,由实例对象构成,是对概念知识量的规定;事件是指概念所属的事件及在事件中扮演的角色,同一概念可能包含在多个事件知识之中,扮演不同的角色,概念扮演的角色不同,拥有的属性也不尽相同;时间和地点描述了概念知识涉及的相关时间和地点信息,概念知识的发布时间和地点有助于揭示概念知识的演化过程;关联是指概念之间的语义联系;来源是指概念知识的文本描述或文献出处。概念知识主题标识由描述概念的语词构成,是对概念主题的揭示。

由表5可知,实例知识表示模型主要揭示了实例知识的属性值、概念、时间、空间、关联和来源。属性值是对实例的具体揭示和描述;概念指出了实例所属的概念类别,同一实例对象可以属于多个不同的概念,在事件知识中扮演不同的角色,从而具备不同的属性和属性值,例如实体对象屠呦呦可以作为“科学家”“女人”等多个概念的实例,作为科学家扮演研究者的角色,作为女人扮演母亲的角色;时间揭示了实例对象存在的具体时间信息;地点描述了实例对象所处的具体地点信息;关联定义了实例对象之间的各种关系;来源是指实例知识的文本描述或文献出处。实例知识主题标识由描述实例的名称构成。

从各种粒度大小知识单元的表示模型可以看出,每个知识单元表示模型除揭示了自身的内部结构(即知识单元所包含的知识元素以及它们之间的关系),以及同一层级的知识单元之间的关系(即相同大小的知识单元之间的关系)之外,还揭示了与其他不同层次的知识单元之间的层级结构(即隶属关系),从而形成了一个多层次的粒结构,所有知识单元层之间的相互关联就构成了全知识空间的整体结构,该结构将不同学科的不同粒度大小的知识单元有序地组织起来,形成一个较为简明的多层次的结构体系。

5 应用案例及讨论

5.1 应用案例

为验证本文提出的数字图书馆馆藏知识资源的跨学科多粒度知识表示模型的适用性和科学性,考虑到图书情报学知识内容具有跨学科特性,广泛涉及认知心理学、管理学、计算机科学、社会学等多个学科领域,笔者选择两篇代表性文献作为实验对象,一篇为来自《中国图书馆学报》的《关联数据发布流程与关键问题研究——以科技文献、科学数据发布为例》[23](下文简称《关联数据发布流程与关键问题研究》),被引频次较高且包含的知识内容较为完善,另一篇为英文文献 *Ontology Integration for Linked Data*[24],来自计算机科学领域期刊 *Journal on Data Semantics*。其中,前者以关联数据为研究对象,探索了关联数据发布问题;后者虽然也是以关联数据为研究对象,但探索了关联数据的本体集成(Ontology Integration)问题。显然,这是两个不同的学科对同一研究对象的不同科研问题进行研究。因此,有关"关联数据"的知识表现出跨学科的特征,具有代表性,能更为直观地展示本文提出的知识表示模型在对数字图书馆馆藏知识资源进行跨学科知识揭示方面的能力。跨学科多粒度知识表示的结果如图 3 所示。

5.2 讨论

通过应用案例(见图 3)分析可以看出,本文构建的数字图书馆馆藏知识资源的跨学科多粒度知识表示模型实现了对数字图书馆各种粒度知识单元的跨学科揭示与描述,不仅可以揭示知识单元的静态特征,而且可以描述知识单元的动态特征,具有一定优势,具体体现在以下几个方面。

(1)多粒度。当前的知识表示方法大多是单粒度的,单粒度的知识表示方法虽然比较简单,但却很难揭示出知识内容的整体框架,难以满足跨学科的知识组织需求,更无法支撑在多个粒度层次上分析问题和解决问题。本文依据知识的抽象程度,对馆藏知识进行了多粒度表示,包括事理知识表示模型、事件知识表示模型、概念知识表示模型和实例知识表示模型四种不同知识粒度的知识表示模型。事理知识最抽象,知识粒度最大,因此事理知识表示模型包含的相关知识元素最多,由事件和事件链构成,如在《关联数据发布流程与关键问题研究》中,事理知识由数据建模、实体命名、实体 RDF 化、实体关联、实体发布和开放查询等事件,以及依据"顺承""组合"关系将事件关联起来的层级事件链构成;事件知识表示模型由概念及其之间的关系等知识要素构成,事件知识是联结事理知识与概念知识的中介,如数据建模事件由定义、实体、RDF 词表、关联等概念构成,这些概念在事件中分别扮演不同的角色,具有紧密的语义关联;概念知识表示模型由概念属性(内涵)和概念实例(外延)构成,是连接事件知识与实例知识的中介,而实例知识是最细粒度的知识单元,实例知识表示模型用于揭示实例对象在特定时空下的具体关系和行为,如"实体"概念包括"期刊""论文"等实例对象,两个实体对象之间存在着"dcterms:isPartOf"关系。

(2)跨学科。现有知识表示模型大都从学科的角度描述知识,学科知识之间的差异给跨学科知识组织造成较大的障碍,然而,千差万别的学科知识又具有共性。首先,本文依据知识的物质统一性规律,在知识表示框架和模型中,融入研究对象和科研问题,以研究对象为中心,以具体科研问题作为学科区分,建立不同学科知识之间的关联,从而在研究对象层次上建立不同学科之间知识的关联。例如,在图 3 中,依据研究对象"关联数据"将两个来自不同学科(图书情报学、计算机科学)的知识关联起来。其次,本文在对各粒度大小的知识单元进行表示时,设置了外部关联属性,通过该属性可以在不同学科之间建立各个知识粒度层次上的语义关联,从而在知识内容层次上打破学科之间的隔阂,实现知识内容的跨学科表示。例如,在图

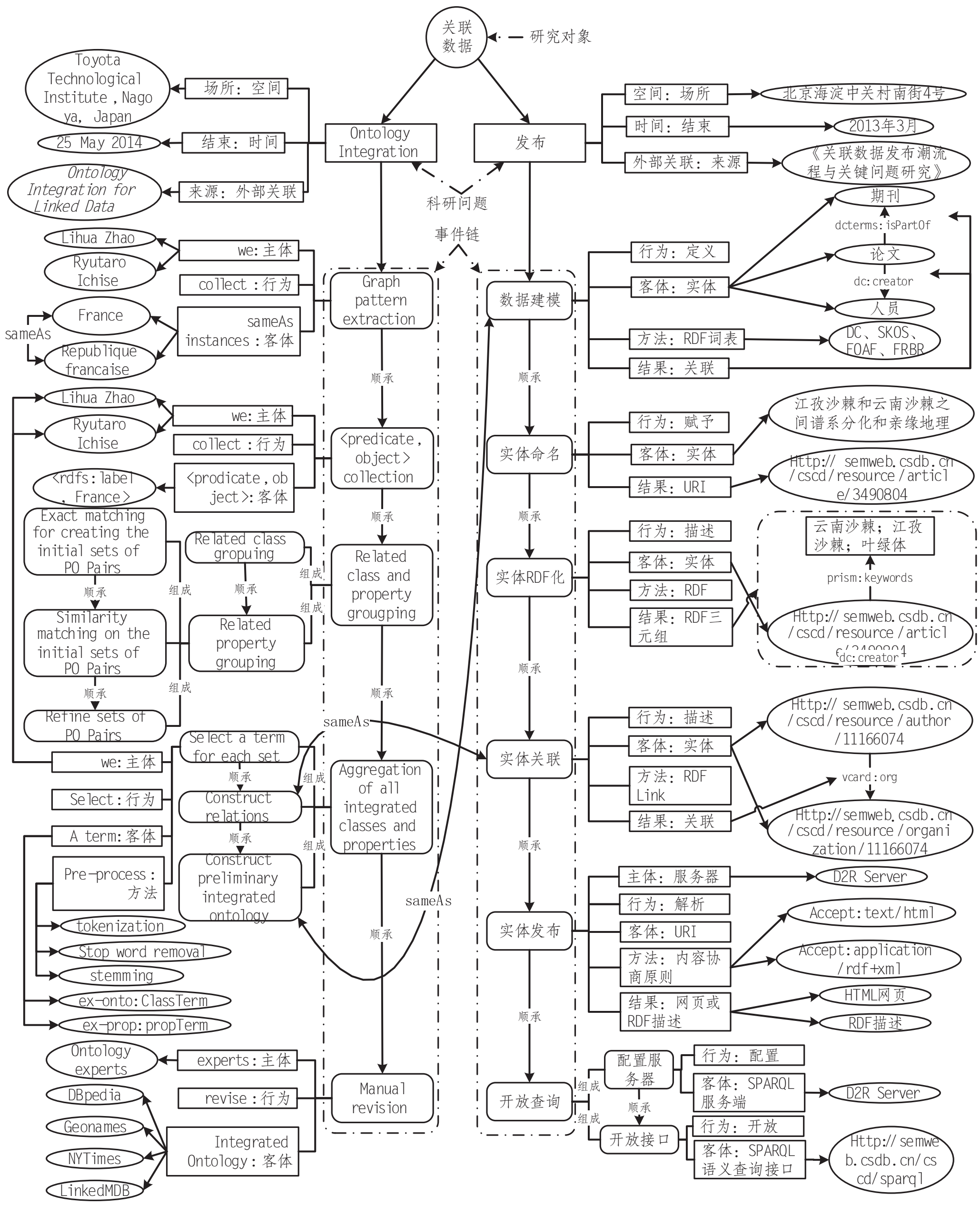

图 3　跨学科多粒度知识表示(部分)

3 中,依据知识单元的相关性将《关联数据发布流程与关键问题研究》中的事件知识“数据建模”“实体关联”分别与 *Ontology Integration for Linked Data* 文中的事件知识“construct preliminary integrated ontology”“construct relations”建立关联,从而实现在不同知识粒度层次上将有关同一研究对象分布在不同学科中的知识聚合在一起。

(3)多角度。现有的知识表示模型大都从

单一的视角描述知识,造成多角度解读知识能力的缺失,影响了多角度分析问题和解决问题的支撑程度。多角度表示某一知识,可以在一定程度上揭示知识的全貌。本文从主题标识、内部属性、情境属性和外部关联属性等多个角度对知识单元进行描述。以文献《关联数据发布流程与关键问题研究》为例,事理知识的主题标识为“发布”,内部属性包括事件和事件链,情境信息为“时间:2013 年 3 月”和“场所:北京海淀中关村南街 4 号”,外部关联属性“出处:《关联数据发布流程与关键问题研究》”;通过对知识资源的多角度描述,最终构建起有关某个知识对象的全景知识视图。

(4)动态性。基于情境描述角色,利用角色描述实体的动态特性。角色与情境密切相关,只有在特定的情境下,角色才能明确,并揭示出各个角色在特定情境下的内在关联。实体对象在不同的情境下扮演不同的角色,具备与特定角色相符的属性和行为,角色的状态、行为等会随着情境的变化而改变。例如在文献 *Ontology Integration for Linked Data* 中的情境为“时间:25 May 2014”和“空间:Toyota Technological Institute, Nagoya, Japan”,它们指出了该研究完成的时间和地点,同时也就决定了“Lihua Zhao”“Ryutro Ichise”此时的角色为科研的“主体”,行为为“关联数据本体集成”的科学研究活动。

(5)语义自然。本文提出的模型遵循人脑对知识的内部表示规律,根据数字图书馆馆藏知识资源自身的特点,围绕研究对象(关联数据)和科研问题(关联数据发布和 linked data ontology integration),由抽象到具体,由粗粒度到细粒度,依据情境信息逐层展开,既保证了文献知识内容的完整表示与语义表达,也实现了知识单元之间逻辑关系的自然揭示与表达。

(6)可扩展。随着科学的发展与研究的深入,知识将不断演化与更新,数字图书馆的馆藏知识资源也会不断累积与变化,这就要求知识表示模型具有可扩展性。本研究提出的数字图书馆跨学科多粒度知识表示模型,利用情境属性来标识知识单元的时空信息,从而在纵向上支撑知识的动态演化扩展;采用抽象和具体化来揭示知识的累积过程,从而在横向上支撑新知识内容的扩展。

6 结论与展望

知识表示是知识组织的基础,知识表示的方式直接决定了上层的知识组织方式。因此,若要实现数字图书馆馆藏知识资源的跨学科知识组织,就需要解决馆藏知识资源的跨学科知识表示问题。然而,跨学科的知识关联主要是建立在文献知识内容层次上的知识单元之间。知识单元具有多粒度性,可以是较为细粒度的实例、数值、概念等,也可以是粗粒度的方案、原理等,因此不同学科的知识之间具有多粒度关联的本质特征,这就要求我们必须深入到文献所载荷的知识内容层次上,对馆藏知识资源进行内容层次上的多粒度知识表示。然而,现有的知识表示方法大都是建立在单一粒度层次上,要么是粗粒度的文献单元层次,要么是细粒度的知识元层次,很少有文献探索知识的跨学科多粒度关联表示问题。

本研究提出的数字图书馆跨学科多粒度知识表示模型从知识内容层次上探索了知识的跨学科多粒度知识表示问题,具有较强的跨学科知识揭示与表达能力。然而,知识表示只是知识组织的第一步,未来还将从以下几个方面继续进行探索:①跨学科知识抽取,若要从知识内容层次对馆藏知识资源进行跨学科知识组织,就需要从文献中抽取出各种粒度大小的知识单元,为跨学科知识组织提供基本的知识对象;②跨学科知识关联,由于学科之间的隔阂,跨学科的知识之间的关联较为稀疏,然而知识在本质上并不存在学科之分,理应拥有大量关联,因此,为实现跨学科的知识组织就需要揭示出这些本该存在的联系,建立不同学科知识之间的

关联;③跨学科知识推理,与单个学科的知识相比,跨学科知识之间存在更多尚未被发现的潜在知识关联,因此,跨学科知识推理就成为跨学科知识组织必须解决的问题。

参考文献

[1] Bliss H E, Dewey J. The organization of knowledge and the system of the sciences [M]. New York: Henry Holtand Company, 1929:356-357.

[2] Szostak R. Classification, interdisciplinarity, and the study of science[J]. Journal of Documentation, 2008,64(3):319-332.

[3] Frodeman R, Klein J T, Mitcham C. The Oxford handbook of interdisciplinarity[M]. Oxford: Oxford University Press, 2010.

[4] Hjørland B, Nissen P K. A substantive theory of classification for information retrieval[J]. Journal of Documentation, 2005, 61(5):582-597.

[5] Ding L, Finin T W, Joshi A, et al. Swoogle: a search and metadata engine for the semantic web[C]// Proceedings of the 13th ACM International Conference on Information and Knowledge Management,2004:652-659.

[6] Zein O K, Kermarrec Y, Garlatti S, et al. A metadata model for web services applied to index and discover e-learning services[C]// Proceedings of the 2nd International Conference on Information & Communication Technologies: from Theory to Applications, 2006:626-630.

[7] Brookers B C. The foundation of information science: part Ⅳ: information science: the changing paradigm[J]. Journal of Information Science, 1981,3(1):3-12.

[8] 裘江南, 王带弟, 仲秋雁. 多相关主题的主题文本资源相关度算法研究[J]. 中国图书馆学报, 2009, 35(4):106-112. (Qiu Jiangnan,Wang Daidi, Zhong Qiuyan. Research on relatedness algorithm between topics and resources of multiple topics[J]. Journal of Library Science in China, 2009, 35(4):106-112.)

[9] Novak J D, Cañas A J. The theory underlying concept maps and how to construct and use them[EB/OL].[2019-01-10].http://cmap.ihmc.us/docs/theory-of-concept-maps.

[10] Quillian M R. Semantic memory [D]. Bedford: Carnegie Institute of Technology,1966.

[11] Minsky M. A framework for representing knowledge[J]. Massachusetts Institute of Technology, 1974(306): 156-189.

[12] Tomkins S S. Script theory: differential magnification of affects[J]. Nebraska Symposium on Motivation, 1978, 26:201-236.

[13] Studer R, Benjamins V R, Fensel D. Knowledge engineering: principles and methods [J]. Data and Knowledge Engineering, 1998, 25(1-2): 161-197.

[14] BernersL T, Hendler J, Lassila O. The semantic web[J]. Scientific American, 2001, 284(5):34-43.

[15] Berners L T. Linked data-design issues [EB/OL]. [2019-01-20]. https://www.w3.org/ DesignIssues/LinkedData.html.

[16] 白海燕,朱礼军.关联数据的自动关联构建研究[J].现代图书情报技术, 2010,26(2):44-49. (Bai Haiyan, Zhu Lijun. Research on automatic interlinking of linked data[J]. New Technology of Library and Information Service, 2010,26(2):44-49.)

[17] 王辉,郁波,洪宇,等.基于知识图谱的Web信息抽取系统[J].计算机工程, 2017,43(6):118-124.(Wang Hui, Yu Bo, Hong Yu, et al. Web information extraction system based on knowledge graph[J]. Computer Engineering, 2017,43(6):118-124.)

[18] 莫曾萌,黄孟洲.形式逻辑[M].成都:电子科技大学出版社,1994:43-86.(Mo Zengmeng, Huang Mengzhou. Formal logic[M].Chengdu: University of Electronic Science and Technology Press,1994:43-86.)

[19] Zacks J M, Tversky B. Event structure in perception and conception[J]. Psychological Bulletin, 2001, 127(1): 3-21.

[20] 张旭洁,刘宗田,刘炜,等. 事件与事件本体模型研究综述[J]. 计算机工程, 2013,39(9):303-307.(Zhang Xujie, Liu Zongtian, Liu Wei, et al. Survey of event and event ontology model research [J]. Computer Engineering, 2013,39(9):303-307.)

[21] Pustejovsky J, Castaño J, Ingria R, et al. TimeML:robust specification of event and temporal expressions in text [C]//Proceedings of the 5th International Workshop on Computational Semantics, 2003:337-353.

[22] Filatova E, Hatzivassiloglou V. Domain-independent detection, extraction, and labeling of atomic events[C]// Proceedings of the 4th International Conference on Recent Advances in Nature Language Processing, 2003: 145-152.

[23] 沈志宏, 刘筱敏, 郭学兵,等. 关联数据发布流程与关键问题研究——以科技文献、科学数据发布为例[J]. 中国图书馆学报, 2013, 39(2):53-62.(Shen Zhihong, Liu Xiaomin, Guo Xuebing,et al. A research on publishing workflow and key issues of linked data: experience with publishing scientific literature and scientific data as linked data[J]. Journal of Library Science in China, 2013, 39(2):53-62.)

[24] Zhao L, Ichise R. Ontology integration for linked data[J]. Journal on Data Semantics, 2014, 3(4):237-254.

王忠义 华中师范大学信息管理学院副教授,硕士生导师。湖北 武汉 430079。

夏立新 华中师范大学信息管理学院教授,博士生导师。湖北 武汉 430079。

李玉海 华中师范大学信息管理学院教授,博士生导师。湖北 武汉 430079。

(收稿日期:2019-04-11;修回日期:2019-06-02)