

知识元的内涵、结构与描述模型研究^{*}

索传军 盖双双

摘 要 新的信息环境下,以文献为单元的知识组织方式已无法满足用户的知识需求,文献中所包含的细粒度知识元成为多学科研究的热点之一。知识元的内涵、结构和描述模型成为亟待研究的重要基础理论问题。本文在对已有知识元概念、分类和描述方法进行总结的基础上,进一步明确了知识元的内涵,并从知识生产的视角,将文献中的知识元划分为常识知识元、引证知识元和创新知识元三种类型,进而从理论和实证分析两方面证明了“一个知识元是 $N(N \geq 1, N$ 为正整数)个语义三元组的逻辑组合”这一假设。此外,为促进知识元的管理和利用,本文借鉴纳米出版思想,构建了知识元的语义描述模型,该知识元描述模型由知识元内容、内容的来源信息和支持信息,以及知识元的出版信息四个要素构成。图 3。表 6。参考文献 52。

关键词 知识元 知识元内涵 知识元分类 知识元结构 知识元描述模型

分类号 G250.2

The Connotation, Structure and Description Model of Knowledge Unit

SUO Chuanjun & GAI Shuangshuang

ABSTRACT

In the new information environment, the way of knowledge organization based on documents cannot meet the users' knowledge needs. In order to achieve fine-grained semantic knowledge organization and refined knowledge service, the fine-grained knowledge units contained in the documents become one of the hot topics in multi-disciplinary research. Revealing the connotation and structure of the knowledge unit becomes an important basic theoretical problem to be studied urgently. Based on the existing concepts, classifications and descriptions, we further clarified the connotation of knowledge unit. From the perspective of knowledge production and creation, the knowledge unit is divided into three types which include common-sense knowledge unit, citation-based knowledge unit and innovative knowledge unit. And then, from the point of theoretic and empirical analysis, we proved the hypothesis of "a knowledge unit is the logical combination of $N(N \geq 1, N$ is a positive integer) semantic triples". In addition, in order to promote the management and use of knowledge unit, we construct the semantic description model of knowledge unit referring to the idea of nanopublication.

Knowledge unit refers to the smallest unit of content that expresses a relatively complete set of specific

^{*} 本文系国家自然科学基金项目“基于关联书目数据的分层聚合与导航机制研究”(编号:71373278)的研究成果之一。(This article is an outcome of the project "Research on Hierarchical Aggregation and Navigation Mechanism Based on Bibliographic Linked Data" (No. 71373278) supported by the National Natural Science Foundation of China.)

通信作者:索传军, Email: suocj@ruc.edu.cn, ORCID: 0000-0002-7416-1531 (Correspondence should be addressed to SUO Chuanjun, Email: suocj@ruc.edu.cn, ORCID: 0000-0002-7416-1531)

knowledge semantically. Common-sense knowledge unit is a generally acknowledged basic knowledge content in public domain and professional subject area. Its citation often does not need to explain or label specific sources and is used to prove the innovative knowledge unit. Citation-based knowledge unit refers to the text content unit that can represent the citation of references. Citation-based units are usually used to demonstrate or support innovative knowledge unit. Corresponding to common-sense unit and citation-based unit, the innovation knowledge unit is innovative professional knowledge, which is a new discovery or new cognition about the specific problem. The innovation knowledge unit can be decomposed into at least one semantic triples of subject, predicate and object. There are logical relations such as coordinate, conditional, explicative, causal, progressive and comparative relations among the N ($N \geq 1$, N is a positive integer) semantic triples of one knowledge unit. Meanwhile, there are also semantic relations such as synonym relation and implication relation.

The description model of knowledge unit consists of content of knowledge unit, provenance and supporting information of content, and publication information of knowledge unit. The content of knowledge unit is a scientific discovery or core scientific conclusion in the document. And it contains three description elements: subject, predicate and object. The publication information is the metadata about the knowledge unit itself. The description elements of publication information mainly include ID, name, creator, creation organization, creation time, regulatory agency and version number of knowledge unit. Supporting information is the background or context information of the content of knowledge unit, which is used to ensure its scientificity and accuracy. Descriptive elements of supporting information mainly include the research method, research data, research object, research tool and theoretical basis for generating innovative knowledge units, as well as the preconditions for the establishment of knowledge unit. Provenance information is the source characteristic information of knowledge unit content, which is used to ensure its traceability, trustworthiness and innovativeness, and to ensure the intellectual property of its creator. The description elements of provenance information mainly include source literature, source journals, source authors, type, language and publishing time.

Due to limited space, this paper selects only one Chinese paper and one English paper in this subject area for case study. In the follow-up, a large sample of examples will be selected to explore the characteristics and structure of knowledge unit to further verify the research hypothesis proposed in this paper. In addition, we only select one example of knowledge unit to verify its descriptive model. 3 figs. 6 tabs. 52 refs.

KEY WORDS

Knowledge unit. Connotation of knowledge unit. Classification of knowledge unit. Structure of knowledge unit. Description model of knowledge unit.

0 引言

新的信息环境下,用户虽然可以更方便地获取所需文献,但却越来越难以找到所需要的知识。目前以文献为单位的知识组织方式,由

于对其包含的知识内容揭示太少,难以为用户提供文献中所包含的知识,因而无法满足用户日益增长的知识需求。针对这一问题,图书情报学、出版学和计算机科学等多个学科领域都在积极探索应对策略。一方面,通过有效的学术评价机制为用户推荐适合他们的高质量文

献,减少用户的文献选择与判断时间;另一方面,直接为用户提供文献中的结论或创新性观点,减少用户的文献阅读时间。事实上,早在20世纪70年代,就有专家指出,知识的控制单元将从文献深化到文献中的数据、公式、事实、结论等细粒度的“知识元”。知识元是知识控制与处理的基本单位,是知识结构的基元。近年来,随着自然语言处理和文本挖掘等相关技术的发展,知识元逐渐成为多学科研究的热点之一。虽然目前关于知识元研究的文献较多,但对知识元的概念、类型和描述方式等还没有形成统一认识,一定程度上影响了以知识元为基础的知识组织和知识服务实践的开展。因而,揭示知识元的内涵、结构与描述模型是亟待研究的重要基础理论问题。

1 知识元相关研究

1.1 知识的基本单元的研究视角

知识是人类认识客观世界过程中形成的智力劳动成果。近年来,随着知识生产速度和规模的增加,用户精细化知识服务的需求日渐增强,在此驱动下,若干学科领域的学者尝试从知识的基本构成单元视角探索人类客观知识的管理和利用。图书情报学和计算机科学是知识构成单元的主要研究领域,教育学和突发事件应急管理领域是主要的应用领域。虽然目前关于知识基本构成单元的形式和概念还没有形成统一认识,但知识元、知识单元、知识因子和知识基因等是主要的代表性观点。

1.1.1 图书情报学的知识组织与知识管理视角

图书馆历来是人类客观知识的传承和服务机构,作为知识服务的基础,现有的以文献为单位的知识组织和管理方式已无法满足用户的知识需求,部分学者转而从知识的基本构成单元视角探索新的知识组织与管理方式。

早在1964年,我国著名文献情报学家袁翰青教授就提出对文献中新发现的知识单元进行组织的思想^[1]。然而,直到21世纪,我国学者

才开始重视以知识的基本构成单元为单位开展知识组织研究。刘植惠^[2-3]在情报基因和思想基因的启发下,提出了知识基因理论,并指出,知识基因是知识继承与发展的最小功能单元,在知识序化过程中具有重要应用。王知津^[4]指出,对知识进行的任何组织都必须建立在知识单元的基础上,而知识单元无非就是概念。王子舟^[5]认为,知识单元是知识的基本组分,是客观知识系统中有实际意义的基本单位,知识组织应该以知识的基本组分为基础。蒋永福^[6]认为,对文献中的知识因子(基本概念)以及知识因子之间的关联关系进行重组,是知识组织的重要方法之一。温有奎^[7]指出,知识元是知识结构的基本组成部分,并提出了基于知识元的知识组织和检索方法。Chang^[8]认为,知识元是相对独立的完整的知识模块,知识元抽取是目前知识组织发展的主要瓶颈,并采用机器学习方法对知识元抽取进行了研究。Zou^[9]和 Lu^[10]等认为,传统知识组织模式是基于资源、主题和元数据目录的,难以表达知识资源中知识元包含的信息内容及其内在关系,因而提出了一种主题—知识元一定位的分布式知识组织管理的新模式。

20世纪末,诞生于知识经济时代的“知识管理”扩展至图书情报领域,成为图书馆学的一个重要研究方向。与“知识组织”相比,“知识管理”的内涵和外延更为宽泛,知识组织可视为知识管理的一个重要环节或目标。虽然不同学者对知识的基本构成单位的理解有所差异,但多数学者强调,知识管理的最小基点应该是知识的基本构成单元。例如,温有奎^[11]指出,研究知识元、知识单元、知识结构之间的信息与知识的变换是知识管理与创新的关键;于秀慧等^[12]构建了基于知识元的知识管理框架;文庭孝^[13]指出,知识单元是进行知识管理的基础,决定了知识管理的广度、深度和精度;黄新红^[14]对知识管理视角下图书馆提供知识单元服务的策略和方法进行了探讨。

1.1.2 计算机科学的知识工程视角

知识工程是研究知识信息处理过程及有关

技术的一门新兴学科,通常被认为是人工智能的一个应用分支。知识工程通过模拟人脑知识存贮模式来研究知识的获取、表示和利用过程,从而设计、构造和维护知识型系统,实现人类问题的自动求解。作为一门以知识为处理对象的分支领域,知识工程主要从粒度化视角研究知识的基本构成单元,并将粒度划分到单个的字、词层面。赵红洲^[15]认为,知识单元是粒子化了的科学概念。陆汝铃等^[16]指出,知识元是采用本体形式表示知识的基本单位。冯琴荣等^[17]基于划分粒度具有定量化表示知识的分类能力,提出了一种知识表示法——划分粒度表示法。徐绪堪^[18]探讨了知识组织中的知识粒度表示和知识粒度化规范。

1.1.3 教育学和突发事件应急管理领域的视角

教育学和突发事件应急管理领域是知识元的重要应用领域。教育学领域的学者认为知识元是组成教学知识点的最小的独立的知识单位,主要研究教材知识元的解析和多媒体知识元库的构建,划分的知识元粒度较大。突发事件应急管理领域主要借助知识元自身的本原性和细粒度性,构建相应的知识元模型和基于知识元的情景库,对应应急管理学科的跨学科性和数据异构性,为应急管理与决策提供支持。

1.2 知识元类型研究

知识元分类是对以文献为基本单元的传统分类体系的细化,是认识、区分知识元的一种逻辑方法,也是提高知识组织和知识管理效率的一种实用方法。从现有研究来看,尽管人们已经认识到基于知识结构的基本单元进行知识管理和知识服务的价值,但目前对知识结构的基本单元还没有形成统一认识,尤其是对知识元的粒度大小以及知识元与知识单元的区别。知识元内涵不统一,导致其类型划分的多样性。

通过文献梳理发现,关于知识元类型的划分,不同学者提出了不同的观点,这些观点主要

集中在图书情报学和教育学领域。目前学者们主要从三个视角对知识元进行分类。①借鉴认知心理学的知识分类思想,根据知识元的表现形式对其进行分类。温有奎^[19]将知识元分成描述型和过程型两大类,其中描述型知识元包含信息报道型、名词解释型、数值型、问题描述型和引文型,过程型知识元包含步骤型、方法型、定义型、原理型和经验型。赵蓉英^[20]在此基础上将智库成果知识元分为陈述型和程序型,前者包括事实知识元、定义知识元、结论知识元等陈述型文字内容,后者包含方法知识元和关系知识元等具有内在结构的文字内容。②根据知识元所表达的内容对其进行分类。毕崇武^[21]将知识元划分为方法型、概念型、事实型和数值型四种类型,张静^[22]根据中小学学科内容将知识元分为概念类、原理类、方法类、事实类和陈述类。③根据知识元在科技文本中的功能,对科技论文所包含的知识元进行分类。秦春秀^[23]将科技文本的内容划分成主题/类别知识元、研究领域知识元、背景知识元、问题知识元、理论/原理知识元等 13 个大类。

通过分析发现,现有的知识元分类方法主要存在以下问题。①知识元的分类标准不统一,类别语言描述不一致,而且不同类别之间存在交叉,不易区分。如文献[19]中的过程型知识元和文献[20]中的程序型知识元,文献[19]中的名词解释型知识元和定义型知识元。②基于知识表达内容和知识元在文本中的功能视角划分的知识元类型多而且细,不利于计算机识别和抽取。如文献[23]和[24]都将知识元划分为 13 种类型。③对知识元内涵的理解存在偏差,混淆了知识元和信息元的区别。如文献[19]中的数值型知识元和文献[20]中的事实知识元其实是文献中所包含的一种信息元,而非知识元。知识元是学术论文中包含的具有一定创新性的,或者具有一定学术和应用价值的知识内容单元,通常传递的是{how/why}问题。信息元是学术论文中所包含的对他人有借鉴作用的信息内容单元,主要指数值型或事实型的具

有完整语义的信息单元,传递的通常是{when/where/who/what}问题。④当前划分的各类型知识元之间缺少一定的逻辑关系或论证关系,如无法体现学术成果的主题性或论证完备性,也无法体现科学知识的继承性和发展性。

1.3 知识元的描述模型研究

知识元描述模型是对知识元的语义内容和结构进行揭示的一种抽象表示,是知识元识别、抽取、标引、组织、存储、关联、重组和再利用的理论基础,其目的是促进知识元的管理和利用。由于对知识元的内涵和分类没有形成统一认识,知识元的描述模型和框架也一直处在发展和完善中。表1呈现了现有文献中的知识元描述模型。

研究发现,知识元描述方法的研究主要集中在图书情报学领域,且具有以下特征。①虽然不同学者给出了不同的描述方法,但平面的线性 N (通常 $N \geq 2$, N 为正整数)元组是知识元的主要表示方式。②从历时角度来看,知识元的描述元素逐渐增加,描述模型日趋完备和全面。如知识元的表示模型从最初的四元组逐渐发展为五元组、六元组和七元组等。③尽管不同知识元模型包含的描述元素不同,但研究发现,这些元素主要都可以归类为内容、属性(内容特征和外部特征)和关系等类别。如名称、类型和主题等元素描述了知识元的内容特征和外部属性特征,导航、来源、消息接口、链接、相关等描述了知识元的关系。

表1 知识元描述模型

作者	年份	描述模型	描述元素类型
温有奎 ^[25]	2003	<名称,属性,操作,导航>	属性、关系
温有奎 ^[7]	2005	<名称,属性,属性描述,消息接口>	属性、关系
温有奎 ^[26]	2005	<名称,表示,法则,操作,导航,上属,相关>	属性、关系
姜永常 ^[27]	2005	<对象名称,对象起因,对象内容,对象过程,对象结论,对象引文>	支持信息、内容
付蕾 ^[28]	2009	<名称,来源,类型,上下文,上下文关系,内容,其他>	属性、内容、关系
温有奎 ^[29]	2011	<名称,属性,值,操作,导航>	属性、关系
姜永常 ^[30]	2011	<名称,属性,操作,链接>	属性、关系
刘森 ^[31]	2012	<编号,导航,来源,类型,特征词内容>	属性、关系
贾生 ^[32]	2012	<编号,名称,内容、主题,类型,来源,其他>	属性、内容
李珊珊 ^[33]	2013	<编号,导航,来源,内容,特征词,知识网络概念表示>	属性、内容、关系
钱飞 ^[24]	2014	<序号,名称,来源,类型,主题词,属性集,内容,上下文,上下文关系>	属性、内容、关系
许春漫 ^[34]	2014	<内容,属性,关系描述信息,历史版本,评论信息,访问记录>	属性、内容
高劲松 ^[35]	2015	<概念,属性,方法,关系>	属性、内容、关系
高国伟 ^[36]	2016	<概念,关系,问题>	内容、关系
毕崇武 ^[21]	2017	<标识组(编码、类别、名称……)、描述组(内容、特征、属性)、关系组(实体关系、来源、出处、责任人……)>	属性、内容、关系
谢庆球 ^[37]	2017	<本我层:编号、来源;自我层:特征词、内容;超我层:内导航、外导航>	属性、内容、关系

现有的知识元描述模型还存在以下几点不足。①多数知识元模型表现为平面的线性形式,缺乏层次性。内容是知识元的核心,其他描述项都是对知识元内容的揭示和描述,而N元组形式的知识元表示方式无法体现这一层次结构。如文献[26]中的描述元素“上属”和“相关”,从其定义看,二者其实均属于知识元的关系对象,文献[28]中的“来源”和“类型”均属于知识元的外部特征属性。②线性的知识元表示方式随着描述元素的增加,容易显得过于冗长,不利于计算机理解和处理。③知识元本身是一个独立的相对完整的内容单元,有其自身的属性信息,如知识元的创建者、创建日期等,现有的知识元模型缺少描述知识元整体的元数据,不利于知识元的管理和利用。

2 知识元的内涵和分类

2.1 知识元的内涵

尽管知识元的概念还没有形成统一认识,但多数学者认为,知识元是构成知识结构的基本单元。我们认为,知识元是指语义上相对完整地表达特定知识的最小的内容单元^[38]。这其中包含四方面的含义:①知识元具有语义相对完整性,即有实际意义和相对独立性;②知识元用于表达特定的知识,如一个科学概念或一条基本原理;③知识元相对于它所表达的特定知识而言,应该是最小的、不可再拆分的;④知识元表现为具体的知识内容。

例如,文献[38]中“绝对创新度是指学术论文所包含的创新知识元与整个科学知识体系或技术体系或科学研究方法体系进行比较,具有多少新发现、新观点、新方法或新技术等”,就是一个关于“绝对创新度”的概念知识元,这一概念知识元揭示了绝对创新度的内涵,具备独立性和语义完整性,而且是不可拆分的,一旦被拆分,则无法完整表达此概念的内涵。

2.2 知识元的分类

任何新的研究成果都是作者在人类已有知

识和他人研究成果的基础上产生的新思想、新方法和新观点。在对现有知识元分类及存在问题进行详细分析的基础上,本文从知识生产和创造的视角,将文献中的知识元划分为常识知识元、引证知识元和创新知识元三种类型。

常识知识元是公共领域和专业学科领域公认的基础性知识内容,其引用往往无需说明或标注具体来源,用于佐证作者所要表达的创新知识元。例如,文献[38]中“老化速度是指学术论文老化的快慢,过去通常用文献半衰期和P指数表示”,是图书馆学领域有关文献老化的一个基本概念,已为图书馆学领域学者所熟知,因而是该文献中的一个常识知识元。

引证知识元是指能够表征施引文献引用参考文献的文本内容,这些文本内容传递的是具体的知识内容,用于论证或支持论文中的创新知识元。例如,文献[38]中“温有奎认为,知识元是指不可再分割的具有完备知识表达的知识单位,是构成知识结构的基元^[15]”,是文献[38]中的一个引证知识元,是对温有奎提出的“知识元”概念内涵的引用。

创新知识元是相对常识知识元和引证知识元而言的创新性专业知识,是作者针对特定问题的新发现或新认识。它是一篇文献中最有价值的知识内容。例如,文献[38]中“学术论文老化速度就是指学术论文中所包含的知识元转移的速度,是单位时间内学术论文所包含知识元转移的数量”,是作者对学术论文老化速度的新认识,因而是该文献中的一个创新知识元。

创新知识元是对常识知识元和引证知识元的继承与发展。事实上,无论是常识知识元、引证知识元还是创新知识元,其在文献中都具体表现为概念、原理、方法、定理、定律、结论等形式。也就是说,常识知识元、引证知识元和创新知识元在本质上是相同的。从“历时”角度看,它们都是不同时期的创新性知识。例如,引证知识元就是被引文献发表之时的创新知识元。当下的创新知识元也会随着发展变为引证知识元和常识知识元。

3 研究假设

由以上论述可以看出,不同学科、不同学者分别从不同角度对知识元进行了一定的分类和描述研究。尽管从名称上看,存在一定的差异,但也达成一些共识,如知识元是构成知识的基本组分,是描述特定领域知识的最小内容单元,是可以被计算机理解的最小语义单元。要使知识元能够被计算机可理解,就需要对知识元进行规范化描述和形式化表达。因而,我们提出假设:一个知识元是 $N(N \geq 1, N$ 为正整数) 个语义三元组的逻辑组合。

该假设包含两个基本问题:①一个知识元可以分解为 $N(N \geq 1, N$ 为正整数) 个语义三元组;②同一个知识元的 $N(N \geq 1, N$ 为正整数) 个语义三元组之间存在一定的逻辑关系。

$$K_U = \{A, R_A\},$$

$$A = \{A_1, A_2, \dots, A_n\},$$

$$R_A = \{R_{12}, R_{13}, \dots, R_{1n}, R_{21}, \dots, R_{2n}, \dots, R_{(n-1)n}\},$$

其中, K_U 表示知识元集合, A 表示一组语义三元组的集合, R_A 表示 n 个语义三元组之间的逻辑关系集合, n 为大于等于 1 的正整数。其中, $R_{(n-1)n}$ 表示三元组 A_{n-1} 与三元组 A_n 之间的逻辑关系, 即 R_{12} 表示三元组 A_1 与三元组 A_2 之间的逻辑关系。另外, 当 $n=1$ 时, $A = \{A_1\}$, $R_A = \{R_{01}\}$, R_{01} 表示该知识元仅有一个语义三元组, 因而, 语义三元组之间的关系不存在, $K_U = \{A_1, R_{01}\} = \{A_1\}$ 。

4 知识元结构分析

科学研究方法包含理论分析和实证分析两种类型, 理论分析是通过逻辑推理的方式对研究问题或研究假设进行描述性分析。实证分析往往采用案例研究、随机抽样调查、实验等具体方法收集资料和数据, 从而得出关于研究问题或研究假设的准确结论。本文分别从理论和实

证两个方面对前文提出的研究假设进行推理和验证。

4.1 理论分析

知识元是指语义上相对完整地表达特定知识的最小的内容单元。内容单元在文献中通常表现为内容相对完整的文本块。因而, 从形式上看, 一个知识元表现为一个文本块。一个文本块由 $S(S \geq 1, S$ 为正整数) 个具有逻辑关系的句子构成。每个句子可以至少用 1 个语义三元组来描述, 即一个句子由 $T(T \geq 1, T$ 为正整数) 个具有逻辑关系的语义三元组构成。语义三元组由主语、宾语以及表达主语和宾语之间语义关系的谓语组成。上述推导可以用以下数学形式来表达:

$$1 \text{ 个知识元} \supset S \text{ 个具有逻辑关系的句子};$$

$$1 \text{ 个句子} \supset T \text{ 个具有逻辑关系的语义三元组};$$

$$1 \text{ 个知识元} \supset S * T \text{ 个具有逻辑关系的语义三元组}。$$

因为 S 和 T 均为大于等于 1 的正整数, 因而, $S * T$ 也是大于等于 1 的正整数, 可以用 N 表示 ($N \geq 1, N$ 为正整数)。所以, 理论上, 一个知识元可以分解为具有逻辑关系的 $N(N \geq 1, N$ 为正整数) 个语义三元组。

4.2 实证分析

文献中的创新知识元是作者关于某研究问题的新发现或新认识, 是在继承与发展常识知识元和引证知识元基础上取得的研究成果。从历史的角度来看, 引证知识元在其被引文献中通常是创新知识元。同样, 常识知识元在其原始文献中也是作为创新知识元被提出。因而, 创新知识元、引证知识元和常识知识元在本质上是相同的。

近年来, 越来越多的学者开始关注学术论文中创新知识点的研究, 这符合知识的生产规律, 而且对加速知识创新具有重要意义。例如, Elsevier 数据库为在线论文提供了亮点 (High-

lights),并将其定义为“论文的核心发现,用以帮助用户快速了解论文的3—5个要点”^[39]。Ronzano等学者^[40-44]借助自然语言处理和机器学习等技术分别从不同视角对学术论文中的新发现和重要结论进行了识别和抽取。

本文在有关创新知识点识别和抽取的研究基础上,基于创新知识元的定义,选取本学科领域中文和英文学术论文各1篇,采用人工判断和抽取的方式,对实例论文中的创新知识元进行结构解析,从实证的角度验证本文所提出的假设,结果见表2和表3。

从实例来看,学术论文中的每一个创新知识元都可以分解为至少1个主语、谓语、宾语形式的语义三元组。知识元所包含的 $N(N \geq 1, N$ 为正整数)个语义三元组通常具有不同的谓语和宾语,但多数情况下共享同一主语,当然也可以有各自不同的主语。由于一个知识元的 $N(N \geq 1, N$ 为正整数)个语义三元组描述的是同一主题的知识内容,因而,这 $N(N \geq 1, N$ 为正整

数)个语义三元组之间必然存在一定的逻辑关系。如表2中概念知识元“学术论文老化速度就是指学术论文所包含知识元转移的速度,是单位时间内学术论文所包含知识元转移的数量”,所包含的两个三元组之间存在并列关系。公式知识元“学术论文含有 m 个知识元和 n 个创新知识元(其中 $m \geq n$)”,则它在学术生命周期

中第 i 年的老化度 $Ai = \frac{\sum_1^i Si}{n}$,创新度 $li = \frac{n - \sum_1^i Si}{m}$ ”所包含的三个语义三元组之间同时存在条件关系和并列关系。

综上,无论理论分析还是实例观察,一个知识元都可以用 $N(N \geq 1, N$ 为正整数)个语义三元组来描述,而且这 $N(N \geq 1, N$ 为正整数)个语义三元组之间存在一定的逻辑关系。因而,一个知识元是 $N(N \geq 1, N$ 为正整数)个语义三元组的逻辑组合这一假设是成立的。

表2 创新知识元的结构解析实例—^[38]

知识元	结构				三元组数量	三元组关系
	三元组	主语	谓语	宾语		
知识元是指语义上相对完整的知识单元。	T1	知识元	是指	语义上相对完整的知识单元	1	
学术论文老化速度就是指学术论文所包含知识元转移的速度,是单位时间内学术论文所包含知识元转移的数量。	T1	学术论文老化速度	是指	学术论文所包含知识元转移的速度	2	T1与T2为并列关系
	T2	学术论文老化速度	是	单位时间内学术论文所包含知识元转移的数量		
学术论文含有 m 个知识元和 n 个创新知识元(其中 $m \geq n$),则它在学术生命周期中第 i 年的老化度表示为 $Ai = \frac{\sum_1^i Si}{n}$,创新度表示为 $li = \frac{n - \sum_1^i Si}{m}$ 。	T1	学术论文	含有	m 个知识元和 n 个创新知识元(其中 $m \geq n$)	3	T1与T2、T3为条件关系;T2与T3为并列关系
	T2	学术论文第 i 年的老化度	表示为	$Ai = \frac{\sum_1^i Si}{n}$		
	T3	学术论文第 i 年的创新度	表示为	$li = \frac{n - \sum_1^i Si}{m}$		

表 3 创新知识元的结构解析实例二^[45]

知识元	结构				三元组 数量	三元组 关系
	三元组	主语	谓语	宾语		
We found that peers' prior use significantly influenced subjects' system use for lower-echelon employees only. (我们发现,同事之前的使用仅对下级员工对系统的使用有显著影响)	T1	peers' prior use (同事之前的使用)	significantly influenced (显著影响)	subjects' system use for lower-echelon employees only (下级员工对系统的使用)	1	
We found that subordinates' prior use influenced subjects' system use for all employees who had subordinates, suggesting a pattern of bottom-up technology diffusion. (我们发现,下属之前对系统的使用会影响有下属的所有员工对系统的使用,表明了一种自下而上的技术扩散模式)	T1	subordinates' prior use (下属之前对系统的使用)	influenced (影响)	subjects' system use for all employees who had subordinates (有下属的所有员工对系统的使用)	2	T1 与 T2 为承接关系
	T2	subordinates' prior use influenced subjects' system use for all employees who had subordinates (下属之前对系统的使用会影响有下属的所有员工对系统的使用)	suggesting (表明)	a pattern of bottom-up technology diffusion (一种自下而上的技术扩散模式)		
Importantly, our findings indicate that social influence patterns differ significantly across groups in an organizational setting; in all likelihood, much depends on the ways in which work is arranged in that organization, and the purpose that the information system serves. (我们发现,同一组织环境下,不同部门的社会影响模式明显不同;原因很可能取决于组织内工作的安排方式和信息系统的服务目的)	T1	social influence patterns (社会影响模式)	differ significantly (有明显差异)	across groups in an organizational setting (在同一组织的不同部门间)	3	T1 与 T2、T3 为解释关系; T2 与 T3 为并列关系
	T2	social influence patterns (社会影响模式)	depends on (取决于)	the ways in which work is arranged in that organization (组织内工作的安排方式)		
	T3	social influence patterns (社会影响模式)	depends on (取决于)	the purpose that the information system serves (信息系统的服务目的)		

5 知识元的描述模型

本文第 4 部分证明了“一个知识元是 N ($N \geq 1, N$ 为正整数)个语义三元组的逻辑组合”这一科学假设。接下来需要考虑的是,如何用形式化的模型描述这一假设,使知识元能够被计算机理解和处理,需要解决两个基本问题:一是语义三元组的描述;二是语义三元组间的逻辑关系及其描述。

5.1 语义三元组的描述

根据前文对创新知识元的结构分析(见表 2 和表 3),一个知识元可以分解为 N ($N \geq 1, N$ 为正整数)个主语—谓语—宾语形式的语义三元组。如创新知识元“学术论文的老化速度就是指学术论文所包含知识元转移的速度,是单位时间内学术论文所包含知识元转移的数量”(见表 2),可以用以下两个语义三元组来描述:学术论文老化速度(主语)→是指(谓语)→学术论文所包含知识元转移的速度(宾语);学术论文老化速度(主语)→是(谓语)→单位时间内学术论文所包含知识元转移的数量(宾语)。

主语—谓语—宾语形式的语义三元组是 W3C 提出的用来描述网络资源的标准数据模型,即资源描述框架 RDF,该模型具有极好的包容性。其中,三元组的主语是被描述的资源。谓语是属性,用来描述资源的某个特定方面——通常是资源的元数据。宾语是属性值,可以是资源也可以是常量。资源包含一切在 Web 上被命名、具有 URI 引用的东西,也可以是无法通过网络和 URI 来定位的对象,如现实中的一本期刊,还可以是非物理存在的抽象概念,如“作者”这个概念。

5.2 语义三元组间的关系类型与描述

本文第 4 部分以创新知识元为例,通过实例分析证明了“一个知识元是 N ($N \geq 1, N$ 为正整数)个语义三元组的逻辑组合”这一假设成立。研究发现,同一个知识元的 N ($N \geq 1, N$ 为正整数)个语义三元组之间存在并列、条件、承接、解释、因果、让步、递进、比较和目的等不同类型的逻辑关系。表 4 列出了同一知识元的 N ($N \geq 1, N$ 为正整数)个语义三元组之间的逻辑关系及其描述示例。

表 4 语义三元组之间的逻辑关系及描述示例^[38]

知识元及其所含三元组实例	逻辑关系	关系描述
①学术论文含有 m 个知识元和 n 个创新知识元(其中 $m \geq n$) ②学术论文第 i 年的老化度表示为 $A_i = \frac{\sum_1^i S_i}{n}$ ③学术论文第 i 年的创新度表示为 $I_i = \frac{n - \sum_1^i S_i}{m}$	①和②为条件关系 ②和③为并列关系	conditionalRelation coordinateRelation
①因为,科学研究的过程是知识在不同科研人员和文献之间转移的过程 ②科学研究的过程部分表现为文献间的引证关系 ③所以,由引文关系构成的引文网络,在一定程度上反映了文献间的知识转移 ④由引文关系构成的引文网络,间接反映了文献作者间的知识转移	①和②为顺承关系 ①、②和③为因果关系	sequenceRelation causalRelation

逻辑关系是根据语义三元组之间的逻辑意义进行划分的。除逻辑关系以外,同一知识元的 $N(N \geq 1, N$ 为正整数)个语义三元组之间还存在一定的语义关系。由于同一知识元的语义三元组之间,以及不同知识元尤其是不同类型知识元的语义三元组之间的关系类型极其复

杂。因而,语义三元组之间的具体关系,以及关系的描述研究将在后续论文中详细论述,不作为本文的讨论重点。表 5 列出了同一知识元的 $N(N \geq 1, N$ 为正整数)个语义三元组之间的语义关系及其描述示例。

表 5 语义三元组之间的语义关系及其描述示例^[38]

知识元及其所含三元组实例	语义关系	关系描述
①学术论文老化速度就是指学术论文所包含知识元转移的速度 ②学术论文老化速度是单位时间内学术论文所包含知识元转移的数量	①和②为同义关系	skos:exactMatch
①知识元是指不可再分割的具有完备知识表达的知识单位 ②知识元是构成知识结构的基元	①和②为蕴含关系	implicationRelation
①知识元是指语义上相对完整的知识单元 ②知识元既可以是一个科学概念 ③知识元也可以是一种新工艺或新方法	①和②、③为例证关系	skos:example
①学术论文含有创新知识元表示有创新性 ②学术论文包含创新知识元数量的多少,反映创新度的高低	①和②为相关关系	skos:semanticRelation

5.3 基于纳米出版物的知识元描述模型

纳米出版物(Nanopublication)由概念网络联盟于2009年提出,是最小的可出版信息单元。纳米出版物将研究结果细分成最小但有意义的科学结论,用基于RDF格式的注释表示这些结论,并将出处信息与结论联系在一起视为一个独立出版物^[46]。简单地讲,纳米出版物就是一个带有相关语境信息的核心科学陈述。当前的纳米出版物框架包含断言、出处和出版信息三个实体^[47-48]。“断言”是一个最小的思想单元,是作者得出的科学事实、实验结果或结论等创新性思想。“断言”实际上就是本文所定义的创新知识元。纳米出版物用一个或多个RDF语义三元组描述文献中的科学结论(创新知识元),这与本文所论证的知识元结构(一个知识元可以分解为N个语义三元组)相一致。“出处”是关于断言的语境的元数据,包括断言的生成方

法以及断言来源于哪里的相关属性。“出处”实体实际上包含“支持信息”和“来源信息”两个子实体。“支持信息”子实体包含断言的研究方法、研究数据、研究对象、研究工具和理论依据等属性;“来源信息”子实体可以用来源文献、来源期刊、作者、机构、基金以及发表时间等属性描述。从前文表1可以发现,现有文献有关知识元的描述模型中,大多数描述元素都可以归类为知识元的内容、属性(包括外部特征和内容特征)和关系三个方面。其中,内容对应纳米出版物的断言,属性和关系则对应纳米出版物的出处信息。“出版信息”是关于一个纳米出版物本身的元数据,包括纳米出版物的创建者、创建日期和版本等^[49]。出版信息可以对知识元整体进行描述,便于知识元的资源化管理,以及方便用户获取和利用。图1是本文参照纳米出版物思想构建的知识元描述模型。与已有描述模型相

比,本文提出的知识元模型层次更加清晰,结构化和语义化更加显著。

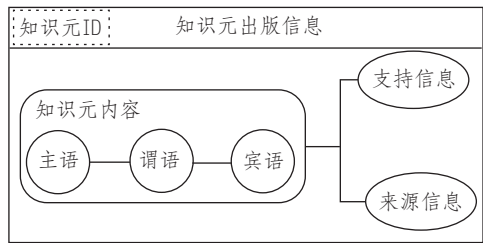


图 1 知识元的描述模型

5.4 知识元描述词表

知识元描述词表就是对描述知识元的相关属性(或元素),包括属性的说明、中英文标签、取值类型和定义域等的规范说明,也就是知识元描述的词汇体系。知识元描述词表的构建应该尽可能复用已有本体中的词汇,如 Dublin Core Terms (DC Terms)^[50]、W3C Provenance Ontology (PROV-O)^[51]、Provenance, Authoring and Versioning ontology (PAV)^[52]等。表6列出了本文构建的知识元的部分描述词汇。

表 6 知识元的描述词表

实体	中文标签	英文标签	取值类型	来源	说明	选自域名空间
(知识元)出版信息	ID	Identifier	字符	系统自定义	指知识元的唯一标识	http://purl.org/dc/terms/
	名称	Title	字符	系统自定义	指知识元的名称	http://purl.org/dc/elements/1.1/
	创建者	wasGenerated-By	字符	系统自定义	知识元创建者可以是团体,也可以是个人	http://www.w3.org/ns/prov#
	创建时间	generatedAt-Time	日期	系统生成	知识元的创建时间	http://www.w3.org/ns/prov#
	责任者	wasAttributedTo	字符	系统自定义	指对创建的知识元的真实性负责责任的机构或个人	http://www.w3.org/ns/prov#
	版本号	hasVersion	数值	系统自定义	指知识元发布的版本号	http://purl.org/dc/terms/
	语种	Language	字符	系统自定义	指知识元发布的语种	http://purl.org/dc/terms/
知识元内容	主语	subject	字符	来源文献	指知识元内容的主语	http://www.w3.org/1999/02/22-rdf-syntax-ns#
	谓语	predicate	字符	来源文献	指知识元内容的谓语	http://www.w3.org/1999/02/22-rdf-syntax-ns#
	宾语	object	字符	来源文献	指知识元内容的宾语	http://www.w3.org/1999/02/22-rdf-syntax-ns#
(知识元内容)支持信息	研究对象	hasResearch-Object	字符	来源文献	知识元内容是在对某实体对象进行研究的基础上获得的	http://ruc.edu.cn/lib/supinfo/
	研究数据	DateCreated	数值	来源文献	指作者得出知识元内容的基础数据	http://purl.org/dc/terms/
	研究方法	Instructional-Method	字符	来源文献	指作者得出知识元内容所采用的研究方法	http://purl.org/dc/terms/

续表

实体	中文标签	英文标签	取值类型	来源	说明	选自域名空间
(知识元内容)支持信息	研究工具	hasResearch-Tool	字符	来源文献	指作者得出知识元内容所采用的研究工具	http://ruc.edu.cn/lib/suppinfo/
	理论依据	hasTheoryBasis	字符	来源文献	指知识元内容的理论依据	http://ruc.edu.cn/lib/suppinfo/
	成立前提	hasPre-conditions	字符	来源文献	指知识元内容成立的前提条件	http://ruc.edu.cn/lib/suppinfo/
(知识元内容)来源信息	创建者	Creator	字符	来源文献	指知识元内容的创建者,对知识元内容的科学性负责的人或组织	http://purl.org/dc/terms/
	来源文献	wasDerived-From	字符	来源文献	指知识元是从何文献中抽取的	http://www.w3.org/ns/prov#
	发表时间	DateIssued	日期	来源文献	指包含知识元的来源文献的发表时间	http://purl.org/dc/elements/1.1/
	类型	type	字符	来源文献	指来源文献的类型	http://www.w3.org/1999/02/22-rdf-syntax-ns#
	国家/地区	createdAt	字符	来源文献	指知识元内容创建者的来源国家或地区	http://purl.org/pav/
	语种	Language	字符	来源文献	指包含知识元内容的来源文献的语种	http://purl.org/dc/terms/
	分类号	Subject	字符	系统自定义	指知识元内容的主题分类号	http://purl.org/dc/terms/

(1)知识元内容是文献中的科学发现或核心科学结论,是知识元描述模型的核心,包含主语、谓语、宾语三个描述元素。

(2)出版信息是关于知识元本身的元数据,明确了知识元的创建者、责任者以及唯一标识等。出版信息的描述元素主要包括知识元ID、知识元的名称、知识元的创建者、创建机构、创建时间、监管机构、版本号以及知识元的引用等。其中,知识元ID是知识元的唯一标识URI,便于知识元的识别、管理、更新与维护。

(3)支持信息是知识元内容的背景或语境信息,用于保证知识元内容的科学性和准确性。支持信息的描述元素主要包括生成创新知识元的研究方法、研究数据、研究对象、研究工具和理论依据等,以及知识元成立的前提条件。

(4)来源信息是知识元内容的来源特征信息,用于确保知识元内容的可追溯性和可信任性,从而保证知识元内容的创新性,以及知识元内容创建者的知识产权。来源信息的描述元素主要包括知识元内容的来源文献、来源期刊、来源作者、类型、语种以及发表时间等。

6 知识元实例的形式化描述

6.1 知识元实例分析

为了验证本文第5部分构建的知识元描述模型的可行性,本文从表2中选取了一个具体的知识元实例进行形式化描述。

实例:学术论文的老化速度就是指学术论文所包含知识元转移的速度,是单位时间内学

术论文所包含知识元转移的数量(见表2)。

(1)知识元内容:通过前文假设可知,一个知识元可以被分解为 $N(N \geq 1, N$ 为正整数)个语义三元组,知识元内容可以用一个或多个RDF三元组进行描述。本例中的知识元内容可以描述为:学术论文老化速度(主语)→是指(谓语)→学术论文所包含知识元转移的速度(宾语);学术论文老化速度(主语)→是(谓语)→单位时间内学术论文所包含知识元转移的数量(宾语)。知识元内容的RDF图具有唯一、可识别的URI。

(2)支持信息:知识元内容的支持信息通常包含多个RDF三元组,三元组主语是知识元内容的URI,宾语是生成知识元内容的研究方法、研究数据和研究对象等,谓语描述主语和宾语之间的关系,如本例中的hasTheoryBasis和InstructionalMethod。

(3)来源信息:知识元内容的来源信息描述方法与支持信息类似,由多个RDF三元组构成,三元组主语是知识元内容的URI,宾语是与知识元内容相关的来源文献、来源期刊、作者和时间等,如本例中的wasDerivedFrom、Creator和type等。

(4)出版信息:知识元的出版信息同样可以用多个RDF三元组来描述,但其主语必须是知识元本身的URI,宾语是知识元的创建者、创建机构和创建时间等,谓语描述主语和宾语之间的关系,如本例中的wasGeneratedBy、generatedAtTime、wasAttributedTo和hasVersion等。

6.2 知识元实例描述

RDF图是由三元组(主语、谓语、宾语)组成的有向图,主语通过谓语指向宾语。本文采用RDF图的形式对6.1中的知识元实例进行描述,结果见图2。

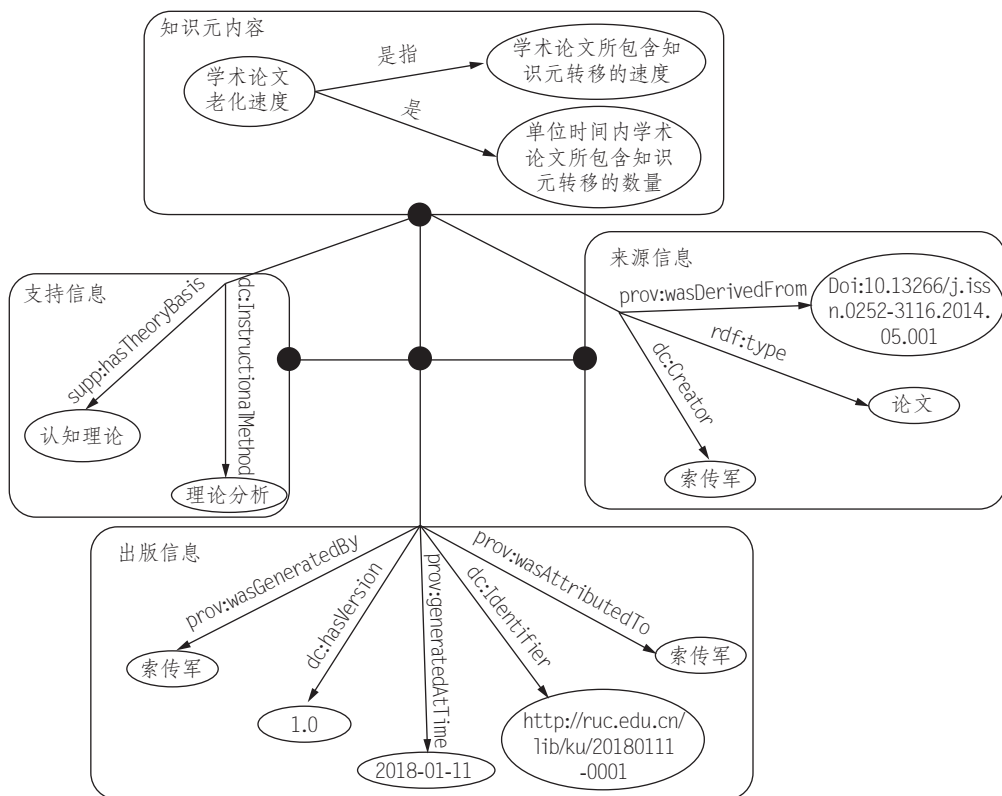


图2 知识元实例描述的RDF图

RDF 与语法无关,它可以建立在不同的语法基础之上。XML 是面向机器阅读和处理的格式。RDF/XML 是 W3C 推荐的标准 RDF 序列化格式,能够在不同应用程序和系统之间进行数

据交换。因而,本文采用 RDF/XML 格式对知识元进行形式化表示,以保证不同数据源之间的信息互操作和计算机自动推理。图 3 是知识元实例的 RDF/XML 表示。

```
@prefix dc:<http://purl.org/dc/terms/>
@prefix dc:<http://purl.org/dc/elements/1.1/>
@prefix prov:<http://www.w3.org/ns/prov#>
@prefix xsd:<http://www.w3.org/2001/XMLSchema#>
@prefix pav:<http://purl.org/pav/>
@prefix rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>

:content {
  rdf:subject:学术论文老化速度
  rdf:predicate:是指
  rdf:object:学术论文所包含知识元转移的速度
  rdf:predicate:是
  rdf:object:单位时间内学术论文所包含知识元转移的数量
}
:provenance {
  :content rdf:type: 论文
  :content prov:wasDerivedFrom:Doi:10.13266/j.issn.0252-3116.2014.05.001
  :content dc:Creator: 索传军
}
:supportingInfo {
  :content dc:Instructional Method: 理论分析
  :content supp:hasTheoryBasis: 认知理论
}
:pubInfo {
  :knowunit dc:Identifier:http://ruc.edu.cn/lib/ku/20180111-0001
  :knowunit prov:wasGeneratedBy: 索传军
  :knowunit prov:wasAttributedTo: 索传军
  :knowunit prov:generatedAtTime "2018-01-11"^^xsd:date
  :knowunit dc:hasVersion: 1.0
}
```

图 3 知识元的 RDF/XML 表示实例

7 结语

本文在对现有知识元定义进行研究的基础上,进一步界定了知识元的内涵,并基于知识生产和创造的视角将文献中的知识元划分为常识

知识元、引证知识元和创新知识元三种类型。本文从理论和实证两个方面对知识元的结构进行了解析,证明了“一个知识元是 $N(N \geq 1, N$ 为正整数)个语义三元组的逻辑组合”这一假设成立。并在此基础上,借鉴纳米出版物思想构建了知识元的语义描述模型。

由于篇幅有限,本文仅选取本学科领域中
文和英文论文各1篇进行实例分析。后续将选
取大样本实例对知识元的特征和结构进行探索,

从而进一步验证本文提出的研究假设。此外,本
文仅选取了一个知识元实例对知识元描述模型
进行了验证,这在后续的研究中有待充实。

参考文献

- [1] 袁翰青. 现代文献工作基本概念[J]. 图书馆, 1964(2): 25-31. (Yuan Hanqing. Basic concepts of modern literature work[J]. Library, 1964(2):25-31.)
- [2] 刘植惠. 知识基因探索(一):第一讲知识基因的涵义及其特征[J]. 情报理论与实践, 1998, 21(1):62-64. (Liu Zhihui. The exploration of knowledge gene (part one): the first talk: the meaning and characteristics of knowledge gene[J]. Information Studies: Theory & Application, 1998, 21(1):62-64.)
- [3] 刘植惠. 知识基因探索(六):第六讲知识基因原理在情报分析研究中的应用[J]. 情报理论与实践, 1998, 21(6):380-382. (Liu Zhihui. The exploration of knowledge gene(part 6): the sixth talk: application of knowledge gene principles in information analysis and research [J]. Information Studies: Theory & Application, 1998, 21(6):380-382.)
- [4] 王知津. 知识组织的研究范围及发展策略[J]. 中国图书馆学报, 1998(4):3-8. (Wang Zhijin. The research scope and development strategy of knowledge organization[J]. Journal of Library Science in China, 1998(4):3-8.)
- [5] 王子舟, 王碧滢. 知识的基本组分——文献单元和知识单元[J]. 中国图书馆学报, 2003, 29(1):5-11. (Wang Zizhou, Wang Biying. The basic components of knowledge - document unit and knowledge unit[J]. Journal of Library Science in China, 2003, 29(1):5-11.)
- [6] 蒋永福, 李景正. 论知识组织方法[J]. 中国图书馆学报, 2001, 27(1):3-7. (Jiang Yongfu, Li Jingzheng. The methods for knowledge organization[J]. Journal of Library Science in China, 2001, 27(1):3-7.)
- [7] 温有奎. 基于“知识元”的知识组织与检索[J]. 计算机工程与应用, 2005, 41(1):55-57. (Wen Youkui. The knowledge organization and searches with the knowledge element [J]. Computer Engineering and Applications, 2005, 41(1):55-57.)
- [8] Chang X, Zheng Q. Knowledge element extraction for knowledge-based learning resources organization[C]//International Conference on Advances in Web Based Learning 2007. [2018-01-25]. https://doi.org/10.1007/978-3-540-78139-4_10.
- [9] Zou J, Liu Q, Yang Z. Knowledge elements mining subsystem of knowledge abstract and fusion system[C]//International Symposium on Knowledge Acquisition and Modeling IEEE, 2008:776-778. DOI: 10.1109/KAM.2008.169.
- [10] Lu H, Feng B, Zhao Y, et al. A new model for distributed knowledge organization management[C]//Seventh International Conference on Grid and Cooperative Computing IEEE Computer Society, 2008:261-265. DOI: 10.1109/GCC.2008.79.
- [11] 温有奎, 赖伯年. 网格技术将带来知识管理革命[J]. 情报资料工作, 2002(5):31-34. (Wen Youkui, Lai Bonian. Grid technology will bring about the revolution of knowledge management[J]. Information and Documentation Work, 2002(5):31-34.)
- [12] 于秀慧, 李宝山. 基于知识元的知识管理[J]. 山东图书馆学刊, 2013(1):10-13. (Yu Xiuhui, Li Baoshan. Knowledge management based on knowledge element[J]. The Library Journal of Shandong, 2013(1):10-13.)

- [13] 文庭孝, 罗贤春, 刘晓英, 等. 知识单元研究述评[J]. 中国图书馆学报, 2011, 37(5):75-86. (Wen Tingxiao, Luo Xianchun, Liu Xiaoying, et al. Review on the research of knowledge unit[J]. Journal of Library Science in China, 2011, 37(5):75-86.)
- [14] 黄新红. 知识管理视角下的图书馆知识单元服务[J]. 湖南科技学院学报, 2008, 29(11):238-240. (Huang Xinhong. Library knowledge unit service from the perspective of knowledge management[J]. Journal of Hunan University of Science and Engineering, 2008, 29(11):238-240.)
- [15] 赵红洲, 蒋国华. 知识单元与指数规律[J]. 科学学与科学技术管理, 1984(9):39-41. (Zhao Hongzhou, Jiang Guohua. Knowledge unit and exponential law[J]. Science of Science and Management of S. & T. 1984(9):39-41.)
- [16] 陆汝钦, 金芝. 从基于知识的软件工程到基于知件的软件工程[J]. 中国科学, 2008, 38(6):843-863. (Lu Ruqian, Jin Zhi. From knowledge-based software engineering to knowware-based software engineering[J]. Science in China, 2008, 38(6):843-863.)
- [17] 冯琴荣, 苗夺谦, 程映, 等. 知识的划分粒度表示法[J]. 模式识别与人工智能, 2009, 22(1):65-69. (Feng Qinrong, Miao Duoqian, Cheng Yi, et al. Knowledge representation using partition granularity[J]. Pattern Recognition and Artificial Intelligence, 2009, 22(1):65-69.)
- [18] 徐绪堪, 房道伟, 蒋勋, 等. 知识组织中知识粒度化表示和规范化研究[J]. 图书情报知识, 2014(6):101-106. (Xu Xukan, Fang Daowei, Jiang Xun, et al. Research on knowledge granularity representation and standardization during knowledge organization[J]. Documentation, Information & Knowledge, 2014(6):101-106.)
- [19] 温有奎, 温浩, 徐端颐, 等. 基于知识元的文本知识标引[J]. 情报学报, 2006, 25(3):282-288. (Wen Youkui, Wen Hao, Xu Duanyi, et al. Text knowledge indexing based on knowledge element[J]. Journal of the China Society for Scientific and Technical Information, 2006, 25(3):282-288.)
- [20] 赵蓉英, 张心源. 基于知识元抽取的中文智库成果描述规则研究[J]. 图书与情报, 2017(1):119-127. (Zhao Rongying, Zhang Xinyuan. On descriptive rules of Chinese think-tank products based on knowledge element extraction[J]. Library & Information, 2017(1):119-127.)
- [21] 毕崇武, 王忠义, 宋红文. 基于知识元的数字图书馆多粒度集成知识服务研究[J]. 图书情报工作, 2017(4):115-122. (Bi Chongwu, Wang Zhongyi, Song Hongwen. Research on the multi-granularity integrated knowledge service based on knowledge element in digital library[J]. Library and Information Service, 2017(4):115-122.)
- [22] 张静, 刘延申, 卫金磊. 论中小学多媒体知识元库的建设[J]. 现代教育技术, 2005, 15(5):68-71. (Zhang Jing, Liu Yanshen, Wei Jinlei. To construct k-12 multimedia knowledge element repository[J]. Modern Educational Technology, 2005, 15(5):68-71.)
- [23] 秦春秀, 刘杰, 刘怀亮, 等. 基于知识元的科技文本内容描述框架研究[J]. 图书情报工作, 2017(10):116-124. (Qin Chunxiu, Liu Jie, Liu Huailiang, et al. The description framework of scientific text content based on knowledge element[J]. Library and Information Service, 2017(10):116-124.)
- [24] 钱飞. 中文文献知识元描述研究[D]. 北京: 中国人民大学, 2014. (Qian Fei. Research on description of knowledge element of Chinese literature[D]. Beijing: Renmin University of China, 2014.)
- [25] 温有奎, 徐国华. 知识元链接理论[J]. 情报学报, 2003, 22(6):665-670. (Wen Youkui, Xu Guohua. Knowledge element linking theory[J]. Journal of the China Society for Scientific and Technical Information, 2003, 22(6):665-670.)

- [26] 温有奎,徐国华,赖伯年.知识元挖掘[M].西安:西安电子科技大学出版社,2005:145.(Wen Youkui, Xu Guohua, Lai Bonian. Knowledge element mining[M]. Xi'an: Xidian University Press, 2005:145.)
- [27] 姜永常.基于知识元的知识仓库构建[J].图书与情报,2005(6):73-74.(Jiang Yongchang. Knowledge warehouse construction based on knowledge element[J]. Library and Information, 2005(6):73-74.)
- [28] 付蕾.知识元标引系统的设计与实现[D].武汉:华中师范大学,2009.(Fu Lei. Design and implementation of knowledge element indexing system[D]. Wuhan: Central China Normal University, 2009.)
- [29] 温有奎.基于知识元的知识发现[M].西安:西安电子科技大学出版社,2011:82.(Wen Youkui. Knowledge discovery based on knowledge element[M]. Xi'an: Xidian University Press, 2011:82.)
- [30] 姜永常.基于知识元语义链接的知识网络构建[J].情报理论与实践,2011,34(5):50-53.(Jiang Yongchang. Knowledge network construction based on knowledge element semantic link[J]. Information Studies: Theory & Application, 2011, 34(5):50-53.)
- [31] 刘淼,王宇.基于主题句的期刊文献知识元库构建[J].情报杂志,2012(11):145-149.(Liu Miao, Wang Yu. Building of periodical literature knowledge element database based on topic sentence[J]. Journal of Intelligence, 2012(11):145-149.)
- [32] 贾生.基于知识元的文献知识仓库构建研究[D].大连:大连理工大学,2012.(Jia Sheng. The research of literature knowledge warehouse construction based on knowledge element[D]. Dalian: Dalian University of Technology, 2012.)
- [33] 李珊珊,王宇.基于HNC理论的期刊文献知识元检索[J].情报杂志,2013(9):190-194.(Li Shanshan, Wang Yu. Retrieval of periodical literature knowledge element based on hierarchical network of concepts theory[J]. Journal of Intelligence, 2013(9):190-194.)
- [34] 许春漫.泛在知识环境下知识元的构建与检索[J].情报理论与实践,2014,37(2):107-111.(Xu Chunman. Construction and retrieval of knowledge element in a general knowledge environment[J]. Information Studies: Theory & Application, 2014, 37(2):107-111.)
- [35] 高劲松,马倩倩,周习曼,等.文献知识元语义链接的图式存储研究[J].情报科学,2015(1):126-131.(Gao Jinsong, Ma Qianqian, Zhou Ximan, et al. Graph storage of semantic literature knowledge element linking[J]. Information Science, 2015(1):126-131.)
- [36] 高国伟,王亚杰,李佳卉,等.基于知识元的知识库架构模型研究[J].情报科学,2016,36(3):37-41.(Gao Guowei, Wang Yajie, Li Jiahui, et al. Knowledge base frame structure research based on knowledge element[J]. Information Science, 2016, 36(3):37-41.)
- [37] 谢庆球,秦春秀,杨智娟,等.知识元层次结构表示方法研究[J].情报理论与实践,2017,40(4):26-31.(Xie Qingqiu, Qin Chunxiu, Yang Zhijuan, et al. Research on hierarchical structure representation of knowledge element[J]. Information Studies: Theory & Application, 2017, 40(4):26-31.)
- [38] 索传军.知识转移视角下的学术论文老化与创新研究[J].图书情报工作,2014,58(5):5-12.(Suo Chuanjun. Study on obsolescence and innovation of academic papers from perspective of knowledge transfer[J]. Library and Information Service, 2014, 58(5):5-12.)
- [39] Elsevier. Highlights[EB/OL]. [2017-12-07]. <https://www.elsevier.com/authors/journal-authors/highlights>.
- [40] Ronzano F, Saggion H. Knowledge extraction and modeling from scientific publications [C]// International Workshop on Semantic, Analytics, Visualization. Springer, Cham, 2016:11-25.
- [41] Dahl T. Contributing to the academic conversation: a study of new knowledge claims in economics and linguistics [J]. Journal of Pragmatics, 2008, 40(7):1184-1201.

- [42] Fisas B, Saggion H, Ronzano F. On the discursive structure of computer graphics research papers[C]// Linguistic Annotation Workshop, 2015:42-51.
- [43] 温有奎, 吴广印. 碎片化科研创新点动态挖掘研究[J]. 数字图书馆论坛, 2014(7):25-32. (Wen Youkui, Wu Guangyin. Dynamic mining of fragmented scientific research innovation points[J]. Digital Library Forum, 2014(7):25-32.)
- [44] 毛琛瑜, 乐小虬. 领域内中文科技文献中新发现语言描述特征分析[J]. 现代图书情报技术, 2016, 32(5):47-55. (Mao Chenyu, Le Xiaoqiu. Linguistic features of new findings in Chinese scientific papers[J]. New Technology of Library and Information Service, 2016, 32(5):47-55.)
- [45] Wang Y, Meister D B, Gray P H. Social influence and knowledge management systems use: evidence from panel data[J]. MIS Quarterly, 2013, 37(1):299-313.
- [46] Chichester C, Gaudet P, Karch O, et al. Querying neXtProt nanopublications and their value for insights on sequence variants and tissue expression[J]. Journal of Web Semantics, 2014(29):3-11.
- [47] Groth P, Gibson A, Velterop J. The anatomy of a nanopublication[J]. Information Services & Use, 2010(30):51-56.
- [48] Concept Web Alliance. What is a nanopublication[EB/OL]. [2017-12-07]. http://nanopub.org/wordpress/?page_id=65.
- [49] Concept Web Alliance. Nanopublication guidelines[EB/OL]. [2017-12-07]. http://nanopub.org/guidelines/working_draft/.
- [50] DCMI Usage Board. Dublin core metadata initiative[EB/OL]. [2018-01-22]. <http://dublincore.org/documents/dcmi-terms/#section-2>.
- [51] World Wide Web Consortium. PROV-Overview: an overview of the PROV family of documents[EB/OL]. [2018-01-22]. <https://www.w3.org/TR/prov-o/>.
- [52] Ciccarese P, Soiland-Reyes S, Belhajjame K, et al. PAV ontology: provenance, authoring and versioning[J]. Journal of Biomedical Semantics, 2013(4):1-22.

索传军 中国人民大学信息资源管理学院教授, 博士生导师。北京 100872。

盖双双 中国人民大学信息资源管理学院博士研究生。北京 100872。

(收稿日期: 2018-01-25; 修回日期: 2018-03-29)