语义出版推动科学论文知识组织:典型模式与应用趋势*

丛天时　郑德俊

摘　要　探讨语义出版及相关技术如何服务于科学论文的知识组织,可以为后续论文文本内容深度体系化、规模化的知识表示与知识关联研究提供参考。本研究在文献调研的基础上明确了语义出版中的知识组织范畴,从纳米出版、微型出版和增强型出版三种典型的语义出版模型特点出发,结合知识组织与语义出版的过程路径,分析了语义出版在知识表示标准化、知识关联关系多维化、知识可重用性和知识更新灵活性等方面的应用优势。未来,可进一步探索多模型融合的领域知识表示方法、生成式人工智能驱动的自动语义标注技术,以及跨文献的可解释语义推理机制,以充分发挥知识组织在科学交流中的整合与调动潜力。图1。表1。参考文献67。

关键词　知识组织　语义出版　纳米出版　微型出版　增强型出版

Semantic Publishing Promotes Scientific Paper Knowledge Organization: Typical Patterns and Application Trends

Cong Tianshi　Zheng Dejun

Abstract: This study aims to explore how semantic publishing and related technologies serve the knowledge organization of scientific papers, providing references for subsequent deep systematic and large-scale research on knowledge representation and knowledge association in scholarly texts. Through literature review, the study clarifies the scope of knowledge organization in semantic publishing, and based on the characteristics of three representative models—NanoPublishing, MicroPublishing, and Enhanced Publishing—analyzes the application advantages of semantic publishing combined with the process and path of knowledge organization and semantic publishing, and from the perspectives of standardized knowledge representation, multidimensional knowledge association, knowledge reusability, and flexibility of knowledge updating. In the future, further exploration can focus on developing domain knowledge representation methods integrating multiple semantic publishing models, intelligent semantic annotation driven by generative artificial intelligence, and interpretable semantic reasoning across publications, so as to fully leverage the integrative and mobilizing potential of knowledge organization in scientific communication. 1 fig. 1 tab. 67 refs.

Keywords: Knowledge Organization; Semantic Publishing; NanoPublishing; MicroPublishing; Enhanced Publishing

科学论文富含海量知识资源,但需要通过复杂的筛选、加工、匹配和分发才能最终服务于用户[1]。在科学论文内容日益增多的背景下,如何高效组织和关联这些知识资源,成为当前亟待解决的重要问题。作为数字出版的高级形态,语义出版正在逐步改变学术界传统的单向输出思维,推动知识服务与出版模式向语义化和智能化方向发展。语义出版提倡采用简洁高效的知识组织形式来表达科学出版物的核心内容和语义价值,在提升知识的可理解性和可重用性的同时,还在提供个性化、定制化的知识服务方面展现了巨大潜力[2]。语义出版通过对数字资源内容进行标注、分割、重组、关联等结构化操作[3],采用特定形式对科学出版物中的知识进行有序化组织,最

* 本文系国家社会科学基金社团项目“面向自主知识创新的索引体系建设与示范应用研究”(项目编号:24SGC080)的研究成果之一。

DOI:10.13666/j.cnki.jnlc.2025.0610

终为用户提供基于语义的非线性阅读体验[4]。在这一过程中,可以将语义出版划分为语义化的知识组织和出版呈现两个主要部分,而知识组织作为语义出版的核心组成部分,承担了整合、关联和呈现知识内容的重要任务。

目前已有部分研究基于出版学视角,从理论和应用层面探讨了语义出版的对象、类型、价值和范畴[5],但仍然缺乏从知识组织理论角度对科学论文语义出版策略的分析。故本研究围绕科学论文知识组织的核心任务——知识表示和知识关联,分析不同语义出版类型中的知识组织策略。通过梳理现有模型与框架,提炼语义出版在科学论文知识组织中的作用,揭示语义出版中知识组织的关键作用,并提出可能的改进策略,为优化语义出版模式和提升科学知识组织效能提供参考。

1 语义出版的知识组织模式与典型形式

1.1 语义出版的研究对象与知识组织范畴

知识组织研究如何对知识进行系统化描述、表示、关联、存储和检索,而语义出版是其在数字学术传播中的典型应用。语义出版的研究对象可以划分为狭义和广义两类范畴,狭义范畴由 David Shotton 于 2009 年首次提出[6],针对载体为学术期刊的科学论文资源,利用知识链接实现语义增强,将论文语义内容和元数据进行关联整合。随着学术信息资源的丰富,语义出版的研究对象也从单一的期刊论文扩展至广义范畴,涵盖图书、学位论文、会议论文、数据集、多媒体内容、社会网络资源[7]乃至科研活动中产生的假设、结论[8]等。现阶段研究仍更多使用狭义的语义出版概念,即主要针对科学论文中出现的多元类型知识的形式、内容和效用[9]开展以语义信息为核心的知识组织。但无论狭义还是广义范畴,语义出版均以知识组织为核心。语义出版与知识组织具有天然的联系,但又并非简单的上下位从属关系或同义替代关系,而是呈现出方法与理论上的逻辑互补。一方面,知识组织理论为语义出版提供了方法论框架,但语义出版并非全盘照搬知识组织的完整体系,不直接涉及分类法、主题法等传统组织工作,而是选择性地采用知识表示、知识关联中与数字出版模式匹配度较高的方法模块。另一方面,语义出版作为知识组织的高级形态,拓展了知识组织的应用边界,促使知识内容分析从科学论文题录数据向全文数据转化。各类语义出版模型作为知识组织理论的实践载体,不仅验证了其方法论的有效性,更在实践中推动了方法创新与服务模式演进,实现了对理论的反哺。其对科学论文知识组织流程的系统优化作用如图 1 所示。

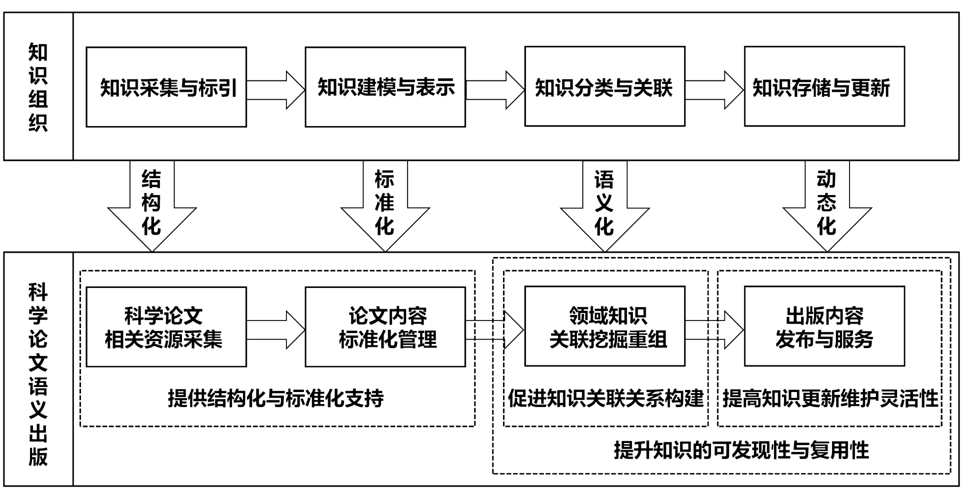

图 1　语义出版驱动下科学论文知识组织的流程优化

语义出版的核心在于通过结构化表示、语义关联和语义知识服务,将传统文献出版内容转化为机器可理解、可计算的知识形式。苏静[10]将语义出版的全流程概括为科学论文相关资源的采集、论文内容的标准化管理、领域知识的关联挖掘重组以及出版内容的发布与服务四个步骤。四部分均涉及了知识组织的研究范畴,对应着知识组织理论中知识采集与标引、知识建模与表示、知识分类与关联、知识存储与更新的完整过程。在知识采集与标引阶段,基于 FAIR 等原则在开放互联网环境下实现多源异构资源的规范化整合,运用 RDF 三元组和领域本体对学术内容进行标准化细粒度语义标注;在知识建模与表示环节,依赖形式多样的出版模型重构 XML/JSON 封装格式下的非结构化、半结构化知识,实现知识的语义化表达;在知识分类与关联层面,通过引文网络、关联数据等构建跨文献、跨模态的动态化语义链接;最终结合网页内容文本增强技术实现内容推荐、交互式阅读、多渠道传播等多方位知识服务[11]。语义出版正在推动传统科学论文出版向智能化、网络化方向演进。

1.2 语义出版的三种典型形式

语义出版以提升科学出版物的可交互性、增强文章关联度为途径,旨在构筑一种新兴的智能化的出版形态,从而更有效地促进学术交流[12]。在知识组织方法的指导下,语义出版已发展出多种实现形式。根据知识单元的粒度、关联强度和服务目标的不同,可归纳为三种特点鲜明且具有代表性的典型模式,分别是纳米出版、微型出版与增强型出版。

(1)纳米出版(NanoPublishing)由概念网络联盟于 2009 年提出[13],是一种将学术研究成果细分为小型知识单元进行发布的出版形式。纳米出版关注科学论文中的研究结论及支持结论的论证,分为内容性和功能性两大部分。内容性部分将科学结论表示为主谓宾形式的 RDF 三元组,并标明结论出处和支持性信息。而功能性部分包括出版物的唯一标识符 ID 和完整性密钥,这二者实际上是纳米出版的真正作用体现,即纳米出版是为计算机存储、索引、访问而生的[14]。知识组织在纳米出版物中更注重对知识进行准确地结构化表示,以便于信息的高效检索与再利用。

(2)微型出版(MicroPublishing)又称微出版,相比于纳米出版关注的内容更为丰富,使用 OWL2 本体语言构建科学论文论证框架[15],涵盖了实体、表述、句子、声明、方法、数据、参考文献等多类型组件单元[16],知识单元间采用多种语义关系进行连接,模块化属性较为明显,且微型出版的支持图谱和反对图谱为后续知识关联提供帮助。知识组织在微型出版中侧重于将文本内容进行分段和不同程度的模块化描述,提供模块间的语义关系,实现跨文段、跨文献的知识关联。

(3)增强型出版(Enhanced Publishing)与纳米出版和微型出版最大的差异在于,增强型出版是一种动态变化的出版形式,旨在通过语义增强技术关联内外部资源,以提高论文内容的可理解性。增强型出版结合了多种信息呈现模态(文本、超链接、图像、视频等),从外部资源链接、内容表达交互和阅读环境样式三方面集成构筑[17]。知识组织在此类出版物中关注如何通过知识链接和上下文信息增强内容的语义关联,实现多模态的知识对齐和知识溯源。

2 科学论文知识组织中的语义出版优势

当前,科学论文在知识组织方面面临诸多挑战。知识碎片化问题导致科学研究内容被细分为大量独立的知识单元,增加了知识组织结构化和标准化的难度[18]。而且论文中知识单元间的关系通常较为复杂,涉及背景、理论、方法、数据、结论等多方关联维度,传统的知识组织方法已经难以清晰地阐释这些内容之间的外在知识关联和内在逻辑表达[19]。同时,不同学科领域的知识体系差异性显著,学科之间的知识组织标准和表示方式各异[20],使得跨学科的知识共享和协同变得困难。随着科学研究的持续发展和知识的动态更新,现有的知识组织体系无法及时回应这些变化[21],也导致知识更新和维护困难,无法满足学术界对最新研究成果的快速获取和利用需求。上述问题严重制约了科学知识的获取、传递与应用,

而语义出版通过其独特的知识组织结构和高度语义化等优势,在知识表示的机器可读性、关联关系的多维构建、知识资源的复用和动态治理等维度实现了方法突破,为解决这些困难提供了切实支持。

2.1 语义出版为知识表示提供结构化与标准化支持

语义出版的目标是实现科学论文中碎片化知识的快速发现、抽取、整合与再利用。而这一目标的实现建立在科学论文本身蕴含的内在组织形式之上,其内容结构天然遵循特定的逻辑框架,通常使用论证结构理论和修辞结构理论[22]来进行知识表达。以图尔敏论证模型[23]为代表的论证结构理论认为,论文的语篇内容实际上是由多个论证单元(论点、论据、论证等)组成的,它们按照一定的逻辑关系推理,形成系统的论证过程。修辞结构理论则更加注重文本功能的层次性和逻辑性,认为语篇由不同功能的文本块(研究问题、方法、结论等)组成,通过修辞关系实现语篇的连贯性[24]。科学篇章修辞块本体 ORB[25]、DoCO 本体模型[26]等典型模型,均强调了修辞结构在科学论文知识组织中的重要性,尤其是文本块之间的功能性关系。知识表示作为知识组织的核心环节,为语义出版提供了形式化表达框架,使得传统线性文本能够转化为结构化、标准化的知识单元。语义出版的核心贡献在于通过形式化方法显示并强化这种内在逻辑,使其从人类可理解的隐含模式转化为机器可处理的显性模式。根据不同的内容表达需求,语义出版也提出了多种用于不同知识粒度[27]和文本结构的知识组织方式[28]。

纳米出版主要基于论证结构理论,应用于细粒度的知识单元独立表达论证内容,将科学论文的知识组织模式由宏观的期刊、文献层次,深化到微观概念的细粒度层次[29]。根据知识表示理论的基本原则,将科学论证陈述从线性表达转换为语义明确的、无歧义的“主体-谓词-客体”RDF 三元组结构,满足知识结构化表示和科学命题逻辑化表达的要求。同时运用领域本体、主题词表约束实现概念消歧、层级划分、规则推理和外部资源语义链接[30,31],并使用固定形式进行内容封装,体现了知识“拆解-关联-整合”的组织路径。纳米出版物需要严格按照技术标准表达面向计算机读写的语义内容,其结构化、标准化程度在三种出版形式中最深。基于这一特点,已有研究使用纳米出版作为一种概念验证方法,规范化表达特定领域概念集[32]、数据集[33]、引文集[34]等内容集合,均取得了较好的概念追溯效果。

微型出版则更倾向于修辞结构理论,应用于中观的模块化知识框架,强调论文中不同模块之间的结构功能连贯性及其修辞关系,为科学论文的语义论证架构提供支撑[35]。微型出版的标准化依赖于定义好的领域本体,Clark 等人最初在生物医学领域提出的微型出版本体只提供了简单的图表、背景、实验结果、实验方法、参考文献以及他们的支持和挑战关系,目的是帮助其他研究人员复现实验[15]。后续研究陆续针对原始实验数据的出处挖掘[36]、药物之间相互作用[37]展开了语义关系更为复杂的微型出版模型构建。这类微型出版本体支持实体、句子、超链接等多种形式的知识集成,是对中等粒度知识的抽象表示和语义组织。

增强型出版则综合了前两者的部分优点,应用于动态的知识增强,直接在网络资源上开展语义关联,通过“根出版物+增强内容”的形式实现分散资源的整合。其技术实现路径为首先采用元数据、本体、URI、Apache Stanbol、大语言模型[38]等结合的方式对根出版物和增强内容进行统一描述。其次根据链路链接、文本相似性、层级隶属关系等进行上下文知识关联建模,实现文献内、跨文献、跨模态内容的精确跳转,从而直接在网络页面上实现增强服务。物理层面万物皆可互联的思想和抽象层面海量知识的多维关联致使增强型出版物的表现形式较为多样,可以分为数据增强、论文整体增强、关联增强和呈现形式增强等[39]。但它必须遵循的重要标准是根出版物及其增强内容必须能够直接定位和访问[40],同时需要符合国际通用数据交换标准[36]。对于科学论文,增强的标准仍是在确保学术内容完整性的前提下,通过结构化、标准化

和可交互的技术手段,实现知识的多维关联。在增强科学论文摘要内容时,需要对摘要进行语步划分,实现结构化处理与语义标注,从而形成清晰一致的摘要增强规范[41]。对学术全文本内容进行增强时需要根据领域本体提取文献内容结构及知识实体,以确保信息抽取的准确性和一致性。

2.2 语义出版促进知识关联多维关系建构

语义出版提供丰富的基于语义的知识共享和关联获取服务[42]。纳米出版和微型出版的知识关联更多依赖 RDF 三元组和本体映射下的科学论文逻辑框架表达和知识内容、知识属性的连结,但增强型出版的知识关联形式、维度都更为复杂,知识不仅仅关联其他知识内容,还可能涉及网页、视频、应用接口等多模态内容。连结这些复杂内容不仅仅需要人工规范下的本体和语义标注,更需要逐渐通过自然语言处理技术实现流程自动化。然而,现有学术领域本体和论证框架因设计目标和应用场景差异,在关系类型定义和覆盖范围上存在差异,亟须通过经典案例的梳理,建立更统一的知识关联理论框架和技术标准。

本文系统整理了国内外语义出版相关的科学论文知识表达与组织体系,将其归纳为 4 大类:语义出版模型、科学论文论证框架、科学论文本体和特殊形式。在此基础上,选取了 8 个具有代表性的模型框架,分别从科学论文模块、知识关联关系、关系跨度、知识粒度和组织方法五个维度进行比较分析(见表 1),主要有以下几点发现:

(1)8 种模型框架均符合论证结构理论或修辞结构理论,几乎涵盖了科学论文的论证全流程,包括题录、正文、参考文献及附录等全文文本信息,但较少涉及多模态的图表内容。其中显性的知识关联以引文关系和论证过程中的修辞关系为主,隐性的知识关联涉及概念的上下位类、知识链接和跨文献的逻辑关系。

(2)模型框架中涉及的知识关联关系较为复杂,多数关联基于抽象概念间预设的逻辑关联(引用、对立、提出、产生等)或人为定义的语义关系(因果、支持、挑战、情境等)。然而,可自动化抽取的关系类型相对有限,主要集中在 RDF 三元组、Web 超链接以及知识图谱等集中组织方法上。

(3)知识关联关系的跨度具有层次性,从微观到宏观可划分为五类:句内关系、段内关系、跨文段关系、跨文献关系及跨载体关系。除科学论文本体局限于单篇论文的范畴外,其余几类模型均突破了上下文的结构约束,表现出较强的跨文献甚至跨载体的关联能力。尤其是语义出版在知识关联维度上展现了强大的拓展能力,关系跨度覆盖了跨文段、文献、载体全模式;关联关系维度也丰富多样,既包括概念间的基础关系,也涵盖实体、引用、逻辑等多维复杂语义关系。

(4)细粒度知识模型更关注知识的微观层次,通常对论文中的具体元素进行详细的语义标注和建模,适合语义出版和科学论文本体等模型,用于构建跨文献甚至跨载体的深层次语义关联。粗粒度知识模型更关注知识的宏观层次,对论文的整体结构和段落进行分类和分析,通过人工编码或结构化的标准描述,支持从段内到跨文段的知识关联,适合于快速理解和组织大范围的学术内容。

(5)8 种模型框架本质上都需要依赖领域本体规范,但自动化的程度有所不同,科学论文论证框架和科学论文本体的自动化程度较低,原因可能是学术领域本体自动化构建较为困难[43]。而语义出版模型由于大量使用语义标注技术,其中自然语言处理、知识图谱等自动化技术都已经较为成熟,故自动化程度较高。Rich HTML 文件作为增强型出版的特殊出版格式,虽然仅涉及知识单元的 Web 链接,不依赖于领域本体,但为了保持语义一致性及解释模块间的复杂关系,仍需要受到本体规范。

综上,上述模型现阶段仍未完全跳出人工定义的关系框架,关系抽象化、概念化程度较高,想要真正实现跨上下文知识内容的语义化关联仍需进一步探索,但已然为科学论文知识关联的理论与实践提供了参考。

表1　科学论文知识表达与组织模型框架

分类	模型	论文模块	知识关联关系	关系跨度	知识粒度	组织方法	参考文献
语义出版模型	纳米出版模型	研究结论及其相关支持信息	概念间关系	句内关系	细粒度	RDF 三元组	[13][14]
	微型出版模型	实体、表述、句子、声明、方法、数据、参考文献	属类关系、表示关系、引用关系、支持关系、挑战关系、情景关系等	跨文段、跨文献	粗粒度	本体	[15][16]
	增强型出版模型	标题、摘要、关键词、引言、方法、结果、结论、附录、参考文献、数据、代码等	引用关系、语义连接、超文本链接	跨文段、跨文献、跨载体	细粒度	本体、知识图谱、语义标注等	[17][39]
科学论文论证框架	研究论文信息结构模式语义框架	研究目标、研究问题、研究假设、研究结果	因果关系、比较关系	段内关系	粗粒度	按标准规范进行人工编码	[44]
	研究设计指纹框架	假说、目标、背景、方法、数据、算法、模型、设备、结果、贡献、结论、未来研究、实体对象	对立、因果、条件、证据、目的等	跨文段、跨文献、	粗粒度	按标准规范进行结构化组织	[45]
科学论文本体	科学论文论证结构本体	前人陈述、问题、目的、假设、方法、结果、结论等	支持、引发、产出、强调、确定程度、证据等	跨文段	细粒度	本体	[46]
	学术论文表述本体	背景、研究目标、研究意义、研究内容、方法、实验、结果、结论等	条件、背景、动机、引出、述评、提出、改进、解决、论证、产生、支持、不支持、总结、目的行为、使用、依据、指导、细化	跨文段	细粒度	本体	[47]
特殊形式	Rich HTML 文件	论文全文、相关链接	页内链接、外延链接、扩展服务链接	跨文段、跨文献、跨载体	细粒度	语义标注、多模态	[48]

2.3　语义出版提升知识的可发现性与复用性

纳米出版、微型出版等语义出版模型都支持对科学论文内容进行颗粒化拆分，将复杂的科学论文拆分为多个独立的知识单元，极大增强了知识的灵活性，提升了知识单元在不同语境下重现并复用的可能[49]。语义出版中大量使用本体模型，本体提供了一个可共享且抽象化的概念模型

用于描述领域知识结构[50],这表明语义出版所关注的就是抽象知识对象及其关联的知识表示统一性和可复用性。已有研究通过分解复杂知识域,使用轻量级的语义模型创建适应动态变化的模块化结构,实现了医学知识图谱的模块化复用[51],提高知识共享和协同的效率。除本体外,增强型出版物由于大量使用结构化的 XML、Rich HTML 技术,通过数字平台间的资源转换和深度标引,形成了规模庞大的主题词表和语义关系群组[52],不仅提高了语义检索能力,同时也大量高效调动了科学论文内可复用的知识资源。最初的科学数据管理的 FAIR 原则就提及了几个与知识的动态更新和维护高度相关的关键词,所有的研究对象都应该具备可重用性(Reusable)、可互操作性(Interoperable)、可发现性(Findable)和可访问性(Accessible)[53],这一原则使得增强型出版提供了大量的可复用元数据和可访问的科学数据[54],并能够与其他来源信息进行复用和融合。

语义出版的任务是通过数据提炼和知识聚合促进期刊文献的知识发现[55]。对科学论文论证结构的语义描述和结构化表达,也是支撑语义出版知识发现的重要途径[56]。语义出版依靠 URI、实体、概念等的链接实现知识关联和知识网络构建,并通过引证、本体、语义谓词[57]等关系将同类内容组织聚合。同时语义出版对文本的语义标注突出了关键概念和语义元数据,为文本添加了价值[58],有助于发现相关概念和文献。Shen 等人就在生物医学领域,通过构建语义框架分析异构信息网络中的语义关联模式,揭示不同学科间的隐藏联系,实现了跨学科的知识发现[59]。

2.4 语义出版提高知识更新与维护中的灵活性

随着知识更新周期的缩短和科学论文数量的急剧攀升,语义出版以知识点进行知识聚类的优势会更为明显[60]。与传统出版形式不同,语义出版中的知识单元以独立、可访问的形式存在,使得知识维护无需重构整个知识体系,仅更新相关知识单元即可。同时利用知识图谱、语义 Web 等技术,整合分散于不同文献、数据库中的知识资源,形成统一的知识网络,当某一资源发生变动时,整个网络可以同步更新。此外,语义出版在基础本体构建时为了保持可复用性,就已经考虑到了本体的版本管理和知识维护。

语义出版对知识更新最大的帮助在于,在 Web 环境下规范了知识的表达和关系的构建。在语义出版中,本体框架为科学论文的知识表示提供了规范,确保了论文内容的统一标引和语义关联[10]。即便是形式更为多样的增强型出版也不是单纯的链接堆叠,而是基于特定的特征和准则[39],进行资源的有序聚合,这使得语义出版在知识表示上具备了高度一致性。同时语义出版还通过精确的知识关联机制,定义了不同科学论文之间的关联关系,如文献引用、论证支持、逻辑关联等,在这些关联模式下新进入 Web 的知识也可以快速自动化地与原有知识产生关联,进一步提升了知识更新和维护的灵活性。

3 基于语义出版的科学论文知识组织提升策略

语义出版的三种代表性形式中,纳米出版规范了细粒度知识单元的表示策略,提升了计算机对语义化知识的理解;微型出版则规范了粗粒度结构要素的表示和组织方法,明确了科学论文应该表征、关联什么语义内容;增强型出版强调全部知识单元在知识网络中的关联和变化。但单一的语义出版形式不足以实现科学论文的智能化、智慧化组织,需将三种表现形式有机结合,推动科学论文粗、细粒度知识的标准化、规模化表示和多维度关联,并逐步向智慧增强出版迈进。

3.1 融合多种语义出版模型的领域知识表示方法

当前多种语义出版模型融合的增强出版方式已初现发展端倪,但其应用效能与成熟度仍表现不足。在国内,截至 2025 年 3 月 1 日,在中国知网中按出版物查询“图书情报与数字图书馆”的核心刊物,仅有两本刊物标注实现了“增强出版”,跳转到期刊官方网站,可以发现其给出了刊物元数据的链接、引文的文内链接、论证结构的导

航目录和较简单的可视化信息,实际上是初步结合了较为基础的纳米出版模型和微型出版模型,而最为核心的语义标注、论证逻辑、语义增强的表达并未体现。国外免费全文数据库中, Pubmed、Elsevier、SpringerLink 实现了更为丰富的论文内图片链接、其他学术外部链接、站内资源导航,其中 Elsevier 实现了部分摘要结构化语义增强(Highlight), SpringerLink 实现了结构化摘要。但整体而言,国内外科学论文的增强型出版形式还停留在初级阶段,深度结合纳米出版的全文语义标注和微型出版的独特论证逻辑表达是语义出版未来需要实现的方向。

语义出版的初衷和现阶段趋势都是实现真正的出版物语义增强,即知识的聚合,而实现知识的聚合无疑需要使用知识组织领域的主题词表、本体等方法技术,这与纳米出版、微型出版的核心技术不谋而合。首先需要运用纳米出版的技术范式对全文进行细粒度语义标注,通过领域主题词表规范知识实体的词形表达,同时结合领域本体建立知识间的语义关系网络。其次需要引入微型出版的论证框架,如采用科学论文结构本体将论文解构为"问题-方法-结论"等结构单元,并通过微型出版建模显式表达结构单元间的逻辑关系和关联程度,形成可计算的论证事理图谱。最终通过增强型出版的动态关联机制,将内部语义标注与外部多模态资源进行链接,形成有层次、有结构、有顺序的"结构化标注-逻辑化组织-动态化关联"完整知识增强链条。

3.2 生成式人工智能驱动的智能化语义标注和语义增强

融合多种语义出版模型的领域知识表示方法实现的困境在于,目前还较少有学术资源数据库网站实现科学论文知识的语义标注,较多停留在关键词语义链接中。其原因是全文语义标注受限于标注类型、规模、上下文信息限制等问题,难以全面普及应用。而现阶段生成式人工智能驱动的语义标注为解决这些问题提供了参考。生成式人工智能的作用不仅在于生成,还在于它背后庞大的可调整的数据支撑使它能够适应几乎所有场景、所有领域的信息抽取和语义标注[61],仅需提供需要标注的内容和提示,就能实现高效准确的自动化标注。大语言模型支持细粒度和多模态的语义组织,已有研究通过 API 调用、参数微调[62]、思维链提示[63]、检索增强[64]等方法提升模型准确性、消除模型幻觉,实现了大规模自动化的科学论文语义元素识别、抽取、分类,相比于传统自然语言处理算法减少了对人工标注的依赖,大大提升了科学论文语义组织的效率[65]。但受限于算力、领域知识的差异性和编辑工作的复杂度[66],生成式人工智能不可能标注所有的科学论文全文,此时就急需智能化、个性化的语义标注服务。一方面,大型的知识服务平台可以从多维度为用户画像,实现精准推送,从而提供针对性的语义标注服务[67]。另一方面,利用增强型出版能够链接外部资源的特性,实现生成式人工智能 API 的实时连接。在用户阅读科学论文前,先使用代价较低的通用模型标注论文内普适性知识;而在用户阅读科学论文时,根据其所选内容智能化提供知识语义解释、外部知识链接、知识图谱关联展示、进一步学习建议等模块。各个学术资源数据库还可以根据其收录科学论文的特征特点,重新训练或微调通用人工智能,以实现更为精准、专业化的语义标注。

3.3 跨文献的可解释语义关系推理

知识的关联也是语义化的一部分,现阶段科学论文论证句内的语义推理已经较为成熟,表达为 RDF 三元组形式,可解释性较强。段内和文内的知识关联可以采用知识元链接、本体等形式进行规范,但由于没有跳出单篇文献范围,仍具备一定可解释性。然而跨文献的知识关联就较为复杂,单纯的共现关系、被引关系等较少关注如何在语义层面深入挖掘文献之间的深层次逻辑,难以真正实现知识的语义推理。

跨文献的可解释语义推理是指利用多个文献中的信息,通过智能化手段,发现文献之间的关系、提取关键的信息,并推导出新的知识或结论。它不仅仅是简单地将不同文献的信息合并,而是要通过语义理解、逻辑推理等手段,将多个文献的知识内容进行综合分析。现阶段跨文献的可解释语义推理集中在引用关系的情感、动机细分和知识

单元的多维关联、网络链接，多是信息的语法和语用层面产生的关联，而对语义信息的使用不充分。跨文献语义推理的核心在于通过语义信息实现不同文献间的知识关联。当前关联路径以人工定义关系规则为主，利用科学论文论证逻辑建立不同论文论证模块的关系承接，从而构建知识推理框架。例如，方法与问题天然存在“解决”与“被解决”的关联，因此跨文献相似的方法与问题之间也可能存在类似联系。然而，现阶段自动化、标准化关联的手段仍显不足，主要依靠语义向量化、层次聚类等传统方法判断跨文献知识的层级隶属关系。这类方法虽然可解释性较强，但需要依赖领域规范化的主题词表，实现较为困难。语义出版的兴起为跨文献知识关联提供了新的技术路径。一方面，通过结构化语义表示和多维语义链接，利用URI统一标识实体、出版模型标准化标注实体，能够实现多元知识的语义归一化，形成统一的概念类并建立跨文献概念链接。另一方面，借助领域本体和科学论文论证本体定义语义规则，可以推理文献中未明示的概念间逻辑关系，同时结合大模型技术解析词语关系，增强推理过程的可解释性。此外，增强型出版更是能够实时链接多论文概念、图表、文献，通过大语言模型溯源原始概念，再从原始概念出发关联更多论文文本，构建动态的知识网络。

结语

语义出版的兴起，标志着科学论文知识组织正经历从传统文献管理向智能化、结构化知识体系构建的深刻转型。在这一范式转换的驱动下，科学论文不再仅是静态的学术记录，而是可以通过语义出版技术转化为动态互联的知识节点，为科学研究的规模化管理、复杂问题拆解和跨学科协作提供基础性支撑。

当前，语义出版的发展仍面临技术融合、标准构建和应用推广的挑战，但其潜力已逐步显现。从线性文本到语义增强，从封闭期刊到开放知识网络，这一演进路径已清晰指明了语义出版的发展方向。面向未来，进一步探索多模型融合的领域知识表示方法、生成式人工智能驱动的自动语义标注技术，以及跨文献的可解释语义推理机制，将有助于提升语义出版在智能知识服务中的知识组织效率与应用价值。可以预见，语义出版将不仅作为学术信息的载体，更将演变为推动科学知识发现、促进跨学科创新和支撑智能科研生态的重要平台，引领科学交流进入一个更加高效、智能和互联的新时代。

参考文献

1 沈锡宾，等.数字化环境下中国科技期刊知识服务模式探析[J].编辑学报，2019(1)：11-16.

2 袁庆，等.中国科技期刊知识服务的实现路径思辨[J].编辑学报，2021(6)：630-634.

3 徐雷.语义出版应用与研究进展[J].出版科学，2016(3)：33-39.

4 徐丽芳，丛挺.数据密集、语义、可视化与互动出版：全球科技出版发展趋势研究[J].出版科学，2012(4)：73-80.

5 苏静，曾建勋.国内外语义出版理论研究述评[J].中国科技期刊研究，2017(1)：33-38.

6 David Shotton. Semantic publishing：the coming revolution in scientific journal publishing[J]. Learned Publishing，2009，22(2)：85-94.

7 Peroni S，Shotton D. FaBiO and CiTO：ontologies for describing bibliographic resources and citations[J]. Journal of Web Semantics，2012，17：33-43.

8 Belhajjame K，et al. Using a suite of ontologies for preserving workflow-centric research objects[J]. Journal of Web Semantics，2015，32：16-42.

9 叶继元.人文社会科学评价体系探讨[J].南京大学学报(哲学.人文科学.社会科学版)，2010(1)：97-110，160.

10 苏静.面向科学交流的语义出版体系建设研究[J].数字图书馆论坛，2018(11)：58-64.

11 华宁，董文杰.数据驱动的增强出版策略及启示——以12种高影响力化学类学术期刊为例[J].中国科技期刊研究，2025(2)：192-198.

12 王晓光，陈孝禹.语义出版：数字时代科学交流系统新模型[J].出版科学，2012(4)：81-86.

13 Mons B，Velterop J. Nano-publication in the e-science era[C]//Workshop on Semantic Web Applications in Scientific Discourse (SWASD 2009)，2009，523：1-4.

14 Giachelle F, et al. Search, access, and explore life science nanopublications on the Web[J]. PeerJ Computer Science, 2021, 7: 335.
15 Clark T, et al. Micropublications: a semantic model for claims, evidence, arguments andannotations in biomedical communications [J]. Journal of Biomedical Semantics, 2014, 5: 1-33.
16 牛丽慧. 微出版及其应用探析[J]. 现代情报, 2018(6): 86-92.
17 宋宁远, 王晓光. 增强型出版物模型比较分析[J]. 中国科技期刊研究, 2017(7): 587-592.
18 苏新宁. 传统知识组织方法的智能力[J]. 科技情报研究, 2024(1): 1-9.
19 牛丽慧, 欧石燕. 科学论文语义标注框架的设计与应用[J]. 情报理论与实践, 2020(3): 124-130.
20 Hjørland B. Theories of knowledge organization— theories of knowledge[J]. Knowledge Organization, 2014, 40(3): 169-181.
21 Helfat C E. Dynamic capabilities and strategic management: organizing for innovation and growth[J]. R&D Management, 2011, 41(2): 217-218.
22 王晓光, 宋宁远. 科学论文内容本体比较研究[J]. 数字图书馆论坛, 2017(8): 2-7.
23 Toulmin S E. The uses of argument [M]. Cambridge: Cambridge University Press, 1958: 21-38.
24 Mann W C, Thompson S A. Rhetorical structure theory: toward a functional theory of text organization[J]. Text-Interdisciplinary Journal for the Study of Discourse, 1988, 8(3): 243-281.
25 马雨萌, 祝忠明. 科学篇章修辞块本体标准及其应用分析[J]. 情报杂志, 2012(10): 112-116, 121.
26 Constantin A, et al. The document components ontology (DoCO) [J]. Semantic Web, 2016, 7(2): 167-181.
27 徐绪堪, 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, 90.
28 徐雷, 叶均玲. 科学交流语义框架模型构建研究[J]. 现代情报, 2022(11): 40-52.
29 吴思竹, 等. 知识资源的语义表示和出版模式研究——以 Nanopublication 为例[J]. 中国图书馆学报, 2013(4): 102-109.
30 Ratnam J, et al. The application of the open pharmacological concepts triple store (open PHACTS) to support drug discovery research [J]. PLoS One, 2014, 9(12): 115460.
31 Fathalla S, et al. SemSur: a core ontology for the semantic representation of research findings [J]. Procedia Computer Science, 2018, 137: 151-162.
32 Williams A J, et al. Open PHACTS: semantic interoperability for drug discovery[J]. Drug Discovery Today, 2012, 17(21-22): 1188-1198.
33 Feijoó M P P, et al. GAP: enhancing semantic interoperability of genomic datasets and provenance through nanopublications[C]//Research Conference on Metadata and Semantics Research. Cham: Springer International Publishing, 2021: 336-348.
34 Fabris E, et al. Nanocitation: complete and interoperable citations of nanopublications[C]//Italian Research Conference on Digital Libraries. Cham: Springer International Publishing, 2020: 182-187.
35 戎军涛, 王丽华. 图书情报领域知识元技术研究述评[J]. 图书馆学研究, 2023(7): 9-15.
36 Gaignard A, et al. From scientific workflow patterns to 5-star linked open data[C]//TaPP'16 Proceedings of the 8th USENIX Conference on Theory and Practice of Provenance, 2016: 44-48.
37 Schneider J, et al. Using the micropublications ontology and the open annotation data model to represent evidence within a drug-drug interaction knowledge base[C]//Workshop on Linked Science 2014—Making Sense Out of Data (LISC2014) at ISWC 2014, 2014: 1-12.
38 Azam S, et al. Exploring technologies for semantic metadata enhancement[C]//International Conference on Complex, Intelligent, and Software Intensive Systems. Cham: Springer Nature Switzerland, 2024: 459-469.
39 李静, 亢小玉. 科技期刊增强出版的融合模式分析与推进策略研究[J]. 编辑学报, 2020(6): 615-618.
40 刘锦宏, 等. 增强型学术期刊出版模式研究[J]. 编辑学报, 2016(1): 15-17.
41 喻琪琛, 王晓光. 科学论文摘要语义增强形式调查研究[J]. 数字图书馆论坛, 2017(8): 8-15.

42 易龙.从数字出版到智能出版:知识封装方式的演进[J].出版科学,2023(1):81-90.
43 李爱华,等.本体构建及应用综述[J].情报理论与实践,2023(11):189-195.
44 Cheng W N,Khoo C S G. Information structures in sociology research papers: modeling cause-effect and comparison relations in research objective and result statements[J]. Journal of the Association for Information Science and Technology,2021,72(11):1367-1385.
45 钱力,等.基于科技文献的研究设计指纹描述框架研究[J].大学图书馆学报,2015(1):14-20.
46 曲佳彬,欧石燕.语义出版驱动的科学论文论证结构语义建模研究[J].现代情报,2021(12):48-59.
47 杜新玉,李宁.学术论文表述本体的构建与应用探索[J].北京信息科技大学学报(自然科学版),2022(6):9-18.
48 Linxi Y U. Analysis on rich HTML application in the highest international impact academic journals of China[J]. Chinese Journal of Scientific and Technical Periodicals,2020,31(6):738.
49 徐雷,潘珺.科学出版物语义数据及其应用研究[J].中国科技期刊研究,2018(7):704-710.
50 Mach M A,Owoc M L. Knowledge granularity and representation of knowledge: towards knowledge grid[C]//Intelligent Information Processing V:6th IFIP TC 12 International Conference, IIP 2010, Manchester, Berlin: Springer, 2010: 251-258.
51 Bravo M,et al. Modularization method to reuse medical knowledge graphs[J]. Applied Sciences,2022,12(22):11816.
52 翁彦琴,彭希珺.爱思唯尔(Elsevier)语义出版模式研究[J].中国科技期刊研究,2014(10):1256-1261.
53 Wilkinson M D,et al. The FAIR guiding principles for scientific data management and stewardship[J]. Scientific Data,2016,3(1):1-9.
54 Jun Z,et al. Linked data and provenance in biological data webs[J]. Briefings in Bioinformatics,2009(2):139-152.
55 徐姝,李玉玲.面向语义出版的图书馆数字人文学术信息服务研究[J].情报探索,2021(9):39-45.
56 王晓光,等.科学论文论证本体设计与标注实验[J].情报学报,2020(9):885-895.
57 郝嘉树.语义出版对图书馆文献资源建设的影响及对策[J].图书馆建设,2016(4):34-38.
58 Shotton D,et al. Adventures in semantic publishing: exemplar semantic enhancements of a research article[J]. PLoS Computational Biology,2009,5(4):1000361.
59 Shen F,Lee Y. Knowledge discovery from biomedical ontologies in cross domains[J]. PloS One,2016,11(8):160005.
60 翁彦琴,等.英国皇家化学会(RSC)——科技期刊语义出版模式的研究[J].中国科技期刊研究,2013(5):825-829.
61 Meyer J G,et al. ChatGPT and large language models in academia: opportunities and challenges[J]. BioData Mining,2023,16(1):20.
62 李西雨,等.基于提示微调的科技论文语义评价指标量化方法研究[J].数据分析与知识发现,2024(Z1):200-212.
63 Wei J,et al. Chain-of-thought prompting elicits reasoning in large language models[C]//Proceedings of the 36th International Conference on Neural Information Processing Systems. Red Hook: Curran Associates Inc,2022:24824-24837.
64 Jiang Z, et al. Active retrieval augmented generation[C]//Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing. 2023:7969-7992.
65 方卿,陈瑞. ChatGPT下科技期刊语义出版的变革[J].图书情报工作,2024(18):4-12.
66 姬建敏,王琪.困境与突围:人工智能赋能学术期刊知识服务的技术架构[J].科技与出版,2023(8):49-55.
67 李静,亢小玉.科技期刊智慧出版模式的构建路径及实施对策[J].编辑学报,2022(5):505-509.

(丛天时　南京农业大学信息管理学院信息资源管理专业2024级博士研究生,郑德俊　教授　南京农业大学信息管理学院)

收稿日期:2025-04-07