

科学论文内容本体比较研究*

王晓光^{1,2}, 宋宁远¹

(1. 武汉大学信息管理学院, 武汉 430072; 2. 武汉大学信息资源研究中心, 武汉 430072)

摘要: 科学论文内容本体规范地定义了构成科学论文的内容组件及其关系, 比较不同本体的差异有助于更全面地理解论文内容单元的多维语义属性, 进而开发更完善的富语义出版物模型。在回顾科学论文内容本体产生及发展历史基础上, 选取具有代表性的6种科学论文内容本体, 从组件及关系角度分析不同本体的差异, 探讨其理论根源和应用。分析表明, 科学论文内容本体类型多样, 但在组件及关系分类和属性定义上缺乏统一的理论框架, 又因领域和情景差异, 在一定程度上制约了内容本体的实例化应用。未来有必要加强用户需求研究, 进一步明确本体开发方向, 提高内容本体的可用性。

关键词: 语义出版; 科学论文; 内容本体; 内容组件; 比较分析

中图分类号: G250.73

DOI: 10.3772/j.issn.1673-2286.2017.08.001

1 引言

为推动知识发现、知识集成等高级知识服务的发展, 提高读者知识获取和科学交流效率, 专业出版领域的语义出版浪潮正在快速兴起^[1]。利用语义技术改变科学知识的组织和关联结构是创新科学论文内容形式的主要路径, 而利用领域和出版本体对科学论文内部多粒度的知识对象和知识单元进行深度语义描述和标注, 是实现科学论文内容网络化、语义化和模块化组织的关键^[2-3], 也是语义出版资源建设层面的主要任务^[4-5]。

科学论文是一类科学文献, 主要由文本和图像两种内容形式组成, 包含表达科学知识的语义单元及其逻辑关系, 如假设、观点、论据、论证过程等。从不同理论视角来看, 科学论文中所蕴含的语义单元和逻辑关系也不相同。为规范地定义科学论文中的语义单元及其关系, 并借其进行科学论文内容的深度语义描述, 人们在不同理论视角下提出不同的科学论文内容本体。在大规模应用这些本体进行实例化建设前, 有必要对这些本体进行比较分析, 以发现其缺陷和不足, 为开发更加系统和实用的内容本体提供参考路线。

2 科学论文内容本体建设的理论基础

科学论文是重要的出版物类型。论文内容本体的设计目标是规范地定义论文的内容组件及其语义关系其设计基础是科学论文内容结构理论。

2.1 科学论文内容结构理论

论文组件是构成科学论文内容结构的基础。修辞结构理论是分析论文组件构成的基础理论, 如被广泛认可的论文结构模型——IMRD模型就是根据修辞结构理论形成的, 该理论将论文的正文部分分成引言、方法、结果和讨论四个模块^[6]。此外, 还有学者提出六模块结构模型^[7], 分别为元信息、定位、方法、结果、解释和成果。ABCDE模型把论文分为注释、背景、贡献、讨论及实体^[8]。在这些研究基础上, 一些更细粒度的内容组件划分框架被提出, 如科学核心概念框架^[9], 定义了假设、动机、目标和客体等单元; 篇章块模型包括研究目标、问题、方法、假设和解释等部分^[10]。

论文内容组件间关系是论文有机整体的关键, 相关理论主要有修辞结构理论和论证理论。修辞结构理论认

* 本研究得到中组部“万人计划”和教育部“新世纪优秀人才”项目资助。

为语篇由功能不同的文本块组成, 根据其价值大小, 这些文本块可以分为核心结构段和辅助结构段, 结构段间存在一定修辞关系以实现语篇连贯, 其关系包括环境关系、阐述关系和证据关系等31种^[11]。论证理论研究成果较丰富, 包括Kunz的问题导向信息系统理论^[12]、Walton的论证框架理论^[13], 以及图尔敏论证模型, 图尔敏认为论证有六个要素, 分别为断言(科学结论)、数据(表示支持断言的基础事实)、保证(解释数据与结论相关的原因)、反驳(表示与保证相关联的例外条件)、支援(用以支持保证的要素)和修饰语(用以表明结论的可信度)^[14]。以上理论是论文内组件关系定义的基础。

2.2 本体理论

本体来源于哲学领域, 说明事物的特征和相互关系, 是“对世界上客观存在物的系统性描述”^[15]。在人工智能领域, Gruber提出了被广泛认可的解释^[16], 其认为“本体是概念模型的明确规范说明”, 具备概念化、明确性、形式化等特点。即本体定义了组成主题领域词汇表的基本术语及其关系, 以及词汇表外延的规则。从组成结构来看, 本体的主要构成部分包括类、属性与实例, 属性又涵盖对象属性和数值属性。其中, 类是对领域概念的规范化说明, 对象属性表示不同类的关系。近年来, 本体工程已逐渐成为知识组织与管理的核心, 本体构建的思想和方法已十分成熟, 如IDEF-5、骨架法等在各个领域均得到广泛运用^[17]。

目前, 在出版领域已提出众多本体, 如面向书目的本体^[18-19]、面向数字出版资源的本体^[20-21]和面向数字出版教学的本体^[22], 并对出版资源和出版活动的相关概念进行了较清晰的定义。在专业出版领域, 对科学论文而言, 清晰地定义其内容组件类型和关系是对论文进行细粒度语义标注和精准知识抽取的基础。为此, Groza等提出标引细粒度文献修辞单元的SALT本体(Semantically Annotated LaTeX)^[23]; 建构出版物内容及引用的语义出版与参考本体集(Semantic Publishing and Referencing Ontology, SPAR本体集)^[24]; 描述生物医学领域内数据组织、论述过程、解释推理的SWAN本体(Semantic Web Applications in Neuro-Medicine)^[25]等。

3 科学论文内容本体

随着语义出版的发展, 越来越多的研究开始关注

如何定义科学论文内容结构。SWAN本体被提出以来, 科学论文内容本体逐渐增多。根据设计目标的不同, 可以将这些本体分为领域型、全域型和专指型三类。

领域型本体通常服务于特定领域的知识表示和知识组织, 将论文内容组件视为领域知识的“容器”, 在描述科学论文语篇的基础上, 注重与领域知识本体的结合。最典型的领域型本体为SWAN, SWAN本体除定义了包括研究声明、研究问题及结构性注解在内的篇章元素外, 还对组织机构、引用及参考文献等进行描述, 同时参考OBO(Open Biomedical Ontologies Foundry)、NCBO(National Center for Biomedical Ontology)等特定学科领域本体的定义。

相较于领域型本体, 全域型本体突破了领域限制, 更加聚焦对科学论文内容组件及其关系的语义表征, 力求全面描述科学论文, 因而粒度更加细致。SALT本体是全域型本体的代表, 主要包括描述结构特征的文档本体、定义修辞功能的修辞本体, 以及起关联作用的标注本体^[23]。

专指型本体则从修辞学、论证理论等不同理论视角出发, 针对科学论文语篇的具体特征进行更细致地定义, 这类本体的针对性更强、关系描述更加准确。代表性本体有修辞块本体(Ontology of Rhetorical Blocks, ORB)^[26]、文献组件本体(Document Component Ontology, DoCO)^[27]、篇章元素本体(Discourse Element Ontology, DEO)^[28]、论证模型本体(Argument Model Ontology, AMO)^[29]等。DoCO、DEO、AMO本体是SPAR本体集的重要组成部分, 且与其他定义出版过程和出版活动的本体共同构成出版本体体系^[24]。

综合已有研究来看, 科学论文内容本体在日渐丰富的同时, 分歧也逐渐明显, 这主要是由于不同内容本体依赖的基础理论不同。科学论文内容结构的划分不唯一, 即使在同一个理论视角下, 因理解不同, 本体中的概念框架也有区别, 这给语义出版实践, 特别是内容组件的自动识别和语义标注造成困惑。为进一步明晰不同本体的差异, 明确不同本体的使用情境, 本文对6种代表性本体(SALT、ORB、DEO、DoCO、AMO和SWAN)从组件类型、关系类型进行细致的比较分析。

3.1 科学论文内容本体组件类型

科学论文内容本体中组件的分类与语言学理论、修辞理论、文献理论有一定关系, 其中修辞理论视角较常用。6种本体的组件类型见表1。

表 1 5种论文内容本体的组件类型比较

本体名称	参考理论与模型	组件模块	粒度
SALT	修辞结构理论/ABCDE模型	摘要、背景、讨论、结论、动机、情景、评价、贡献、实体	粗粒度
ORB	修辞结构理论/IMRD模型	头部(题目、作者、机构、出版地), 主体(引言、方法、结果、讨论), 尾部(致谢、参考文献)	粗粒度
DEO	修辞结构理论/ABCDE及IMRD模型	致谢、作者贡献、背景、参考文献、个人简介、图注、结论、数据、数据集描述、题词、讨论、结语、评价、未来工作、引言、图例、材料、方法、模型、动机、附录、问题描述、序言、参考文献相关工作、结果情景、补充信息	细粒度
DoCO	修辞结构理论/ABCDE及IMRD模型	修辞功能(摘要、附录、前言、后记、参考书目、序言), 结构特征(引用块、章节、图表、表格、段落、文本块、列表、部分、标题)	细粒度
AMO	论证理论/图尔敏模型	断言、证据、支援、反驳、保障、修饰语	细粒度

SALT本体主要借助论文的修辞结构定义内容组件。该本体在ABCDE模型基础上进行扩充, 确定了组件模块包括摘要、动机和贡献等。ORB是W3C组织提出的推荐标准, 包括头部、主体和尾部。头部主要是对出版物附加信息的描述, 包括题目、作者、机构及出版地等条目; 主体则借鉴IMRD结构, 将正文部分分为引言、方法、结果及讨论; 尾部包括致谢和参考文献。相对而言, ORB本体对论文内容组件定义的粒度较粗, 所以适用面更广。DEO主要从科学论文内容组件的修辞功能出发, 定义了包括致谢、背景和方法等多种具有不同修辞功能的组件。DoCO对DEO进行了补充, 采用OWL2作为描述语言, 从结构特征和修辞功能角度定义文献组件。

科学论文内容本体中对论文组件的定义经历了由简单到复杂的过程。较早的SALT本体定义了较完整的论文组件框架, 但这些组件定义的粒度较粗, 不能更细致地揭示科学论文各部分的属性。DEO和DoCO借鉴SALT本体的分类体系, 更细致、全面地定义了论文组件, 包含的类别更加丰富。同时, 这些本体开始关注新兴的内容部分, 如关于数据集的描述和关于外部资源的描述等。AMO本体对论文内容组件的定义与SALT、ORB、DEO、DoCO不同, 其主要从论证的视角对内容部分进行划分, 这种划分有助于解析论文的论证单元和论证结构, 构建跨篇章的论证网络。

综上, 现有论文内容本体以修辞理论为基础, 其目标是提供机器可读的论文组件框架^[30]。随着论文内容本体的发展, 关于组件的定义越来越细致、全面。

3.2 科学论文内容本体关系类型

目前, 大多数科学论文内容本体对组件间关系的描述都借助对象属性来定义, 对象属性通过关联不同类, 实现文献语义结构及组件关系的显性化, 以适应计算环境下的知识计算。下文对SALT、SWAN、AMO^[31]本体中定义的关系类型进行比较分析(见表2)。

表 2 3种论文内容本体中关系类型的比较

本体名称	理论基础	关系
SALT	修辞结构理论	判断、证据、环境、对照、让步、目的、重述、条件、解释
SWAN	-	参考、相关、回应、动机、引发、替代
AMO	论证理论	证据支援、断言、引发、证明、支持等

SALT本体在修辞组件基础上, 定义了9种关系, 包括判断、证据、环境等。与其类似的还有Naja-jazzar等提出的OntoReST本体^[32], 也定义了论文组件间的9种关系, 包括背景、对比、解释等。AMO本体以论证理论为基础, 针对不同的论证组件类型, 界定了支持、证明和涉及等21种组件关系。SWAN本体对科学论文的篇章关系进行了定义, 在经历三次修订后, 借助引文类型本体(Citation Typing Ontology, CiTO)实现关系类型的完善, 并具备一定层级结构, 主要关系包括指代、相关和替代等。

现有本体对科学论文组件间关系的定义较简单。修辞结构视角下的关系定义侧重于解释论文的文本组织结构, 论证理论视角下的关系侧重于解释观点的组织。

在一定程度上, 论证关系可被视为特殊的修辞关系, 用来表示不同观点间的逻辑关系。相较于修辞关系, 在科学论文语境下, 论证关系对不同知识点逻辑关系的表达能力更强, 更加符合科学论文的知识表示要求, 因此成为组件间关系的主流描述方案。尽管定义方式不同, 但基本可归纳为证明、支持、同意和建议四类, 并可根椐情感倾向的不同进行更细致的定义。

4 科学论文内容本体的应用

准确和规范地定义科学论文的组件及其关系是实现语义出版的基础工作之一, 科学论文内容本体起到关键作用。

目前, 对科学论文进行深度语义标注是科学论文内容本体最直接的应用。相关工作包括Ciancarini等设计语义棱镜模型, 通过AMO、DoCO等科学论文内容本体的应用, 从不同角度实现对科学文献内容及情境的语义标注^[33]; Bartalesi等利用DoCO、FaBiO等本体以及FRBR框架, 构建拉丁文学作品结构框架, 同时利用SKOS、CIDOC-CRM等标准进行文本语义标注实验^[34]。科学论文内容结构复杂, 单一本体通常不能全面揭示科学论文内容组件。因此, 在进行语义标注时, 一般使用多个本体从不同角度对文献对象及其知识内容进行语义描述。图1表示利用科学论文内容本体对科学论文的不同部分进行语义标注。

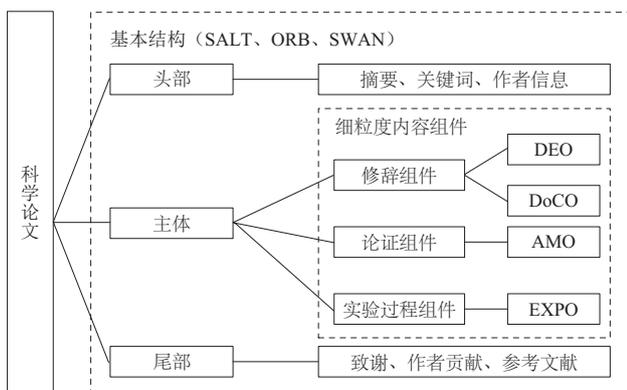


图1 科学论文内容本体的应用

本文在图1的基础上, 综合运用SALT、ORB、DEO等本体对文献进行标注实验, 首先使用ORB对论文修辞框架进行标注, 针对各部分细粒度的文献内容组件利用DEO、SALT等本体进行细粒度深度语义标注。在此以*A Study of Lexical Distribution in Citation*

Contexts Through the IMRaD Standard^[35]为例, 对论文内容组件进行表示。经过标注实验可以发现, 同时使用不同科学论文内容本体, 可以较全面地描述与表示科学论文内容组件。

借助科学论文内容本体可以对科学论文进行深度的语义标注, 进而实现科学论文的细粒度内容抽取与语义关联。在内容抽取方面, Sateli等以SALT和DoCO等本体为基础构建一系列抽取规则及流程, 进行断言和贡献部分的自动抽取实验, 提出将科学文献转换成语义知识库的方法^[36]。Malhotra等借助机器学习, 结合SWAN本体, 构建HypothesisFinder, 用以从科学论文中抽取假设构建知识库^[37]。在语义关联方面, Utopia Documents项目使用DoCO本体对PDF文档进行标注, 作为实现外部知识自动关联、可视化展示及动态交互的基础^[38]; 该项目还发布Web客户端PDFX^[39], 将PDF文档解构为半结构化XML文档后, 借助DoCO本体进行语义标注。Biotea项目也以DoCO本体为基础, 结合基因本体和医学主题词表等, 在对科技文献结构进行定义的同时, 设计一套基于RDF的文献与知识表示模型, 对PubMed中的文献进行语义关联^[40]。SLOR (Semantic Linkage Open Repository) 项目使用CiTO、DoCO等本体, 借助引用关系实现文献信息对象间关联^[41]。

5 现有科学论文内容本体的不足

通过以上比较分析和应用情况调查, 可以发现科学论文内容本体的设计仍存在一定的不足, 大体上可以分为三个方面。

5.1 内容本体设计的出发点有待调整

需求是设计人工产物的基础, 本体作为人工产物, 也必须满足一定的需求。内容本体是一种形式表达本体, 目的是描述论文的内容组成与结构。以此本体进行论文内容的语义标注, 可以实现内容的重组和挖掘。目前, 大部分论文内容本体中的组件设计是以修辞理论为基础, 而修辞理论主要用于人工阅读, 非机器阅读, 这意味着以该本体进行的论文内容语义组织, 不是以支持机器读取和“理解”内容为目标, 所以难于支撑后期的知识抽取、集成等内容再组织需求。为实现论文内容的再组织, 还需要考虑内容片段的语用价值, 而非修辞意义, 所以未来还需据此开发新型的论文内容本体。

5.2 内容本体设计需考虑学科差异

科学论文内容的领域性突出,如医学领域与人文领域论文内容结构差别较大,论文内容的论述方式也有较大差别。论文内容本体若要发挥实用价值,还必须考虑这些差别,因此有必要对内容本体进行分层,提出多领域共享的高层概念和各领域独有概念,以此提高本体的实用性。

5.3 内容本体设计需考虑新的内容组件形式

科学论文是科学知识的容器,最基本的表现形式是字符构成的文本以及各种形式的图表。随着数字出版的发展,一些新型的论文内容组件形式逐渐浮现,如可执行组件、音频视频、互动表格、数据集等,且这些内容模块具有独特作用。如何对这些组件,甚至是组件内部的子件部分进行定义和表示,仍需内容本体设计者做进一步研究,考虑如何扩展内容本体的属性和关系设计。

6 总结

科学论文既是知识的容器,也是科学交流的主要载体,随着网络出版和“第四研究范式”的兴起,语义出版逐渐成为科学交流系统和知识组织交叉研究的前沿领域。利用科学论文内容本体对特定的知识单元进行语义标注,是实现科学论文内容从叙事化的文本向结构化的、支持语义计算的数据过渡的基础。科学论文内容本体在此过程中扮演着不可或缺的角色。

为分析现有内容本体的差别及其发展趋势,明确未来论文内容本体设计的方向,本文对6种论文内容本体中的组件和关系定义进行了定性的比较分析,总结其共性与差异,并在此基础上对现有内容本体的不足和未来发展方向进行论述。这些工作为开发更完善的科学论文内容本体以及语义出版实践提供了指导。

参考文献

- [1] SHOTTON D.Semantic publishing:the coming revolution in scientific journal publishing[J].Learned Publishing,2009,22(2):85-94.
- [2] KIRCZ J G.New practices for electronic publishing 2:new forms of the scientific paper[J].Learned Publishing,2001,15(1):27-32.
- [3] THANOS C.A vision for open cyber-scholarly infrastructures[J]. Publications,2016,4(2):13.
- [4] 王晓光,陈孝禹.语义出版的概念与形式[J].出版发行研究,2011(11):54-58.
- [5] 李楠,孙济庆,马卓.面向学术文献的语义出版技术研究[J].出版科学,2015,23(6):85-92.
- [6] BURROUGH-BOENIS J.International reading strategies for IMRD articles[J].Written Communication,1999,16(3):296-316.
- [7] HARMSZE F A P.A modular structure for scientific articles in an electronic environment[D].Amsterdam:Universiteit van Amsterdam,2000.
- [8] De WAARD A,BREURE L,KIRCZ J G,et al.Modeling rhetoric in scientific publications[C]//Proceedings of the International Conference on Multidisciplinary Information Sciences and Technologies.InSciT2006,2006:25-28.
- [9] LIAKATA M,TEUFEL S,SIDDHARTHAN A,et al.Corpora for the conceptualisation and zoning of scientific papers[C]//International Conference on Language Resources and Evaluation.2010,41(4):105-108.
- [10] DE WAARD A,BUITELAAR P,EIGNER T.Identifying the epistemic value of discourse segments in biology texts[C]//Proceedings of the Eighth International Conference on Computational Semantics.[S.l.]: Association for Computational Linguistics,2009:351-354.
- [11] 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.
- [12] KUNZ W,RITTEL H W J.Issues as elements of information systems[M]. Berkeley:Institute of Urban and Regional Development,University of California,1970:5-10.
- [13] WALTON D,REED C,MACAGNO F.Argumentation schemes[M]. London:Cambridge University Press,2008:3-23.
- [14] TOULMIN S E.The uses of argument[M].London:Cambridge University Press,2003:5-20.
- [15] 邓志鸿,唐世渭,张铭,等.Ontology研究综述[J].北京大学学报(自然科学版),2002,38(5):730-738.
- [16] GRUBER T R.A translation approach to portable ontology specifications[J]. Knowledge Acquisition,1993,5(2):199-220.
- [17] 杜小勇,李曼,王大治.语义Web与本体研究综述[J].计算机应用,2004,24(10):14-16.
- [18] 王文清,刘春彤,张月祥,等.PUBO:面向出版的数字资源本体建模[J]. 大学图书馆学报,2015,33(3):88-95.
- [19] 阮怀伟.面向数字化出版的教育资源本体建模[J].电脑与信息技术,2013,21(5):52-54.
- [20] 司莉,陈雨雪,庄晓喆.基于主题词表的数字出版领域本体构建[J]. 出版科学,2015,23(6):80-84.
- [21] D' ARCUS B.Bibliographic Ontology Specification[EB/OL]. [2017-02-14].http://bibliographic-ontology.org.

- [22] SwetoDblp Ontology[EB/OL].[2017-02-14] <http://lsdis.cs.uga.edu/projects/semdis/swetodblp/>.
- [23] GROZA T,MÖLLER K,HANDSCHUH S,et al.SALT:weaving the claim web[C]//International the Semantic Web & Asian Conference on Asian Semantic Web.Berlin:Springer,2007.
- [24] PERONI S.Semantic publishing:issues,solutions and new trends in scholarly publishing within the semantic web era[D].Bologna: University of Bologna,2012.
- [25] CICCARESE P, WU E, WON G,et al.The SWAN biomedical discourse-ontology[J].Journal of Biomedical Informatics,2008,41(5):739-751.
- [26] 马雨萌,祝志明.科学篇章修辞块本体标准及其应用分析[J].情报杂志, 2012,31(10):112-116.
- [27] The Document Component Ontology[EB/OL].[2016-09-16].<http://www.sparontologies.net/ontologies/doco/source.html>.
- [28] The Discourse Element Ontology[EB/OL].[2016-09-15].<http://www.sparontologies.net/ontologies/deo/source.html>.
- [29] VITALI F,PERONI S.The Argument Model Ontology(AMO)[EB/OL].[2016-09-15].<http://www.essepuntato.it/2011/02/argumentmodel>.
- [30] CONSTANTIN A,PERONI S,PETTIFER S,et al.The document components ontology(DoCO)[J].Semantic Web,2016,7(2):167-181.
- [31] CICCARESE P,SHOTTON D,PERONI S,et al.CiTO+SWAN: the web semantics of bibliographic records,citations,evidence and discourse relationships[J].Semantic Web,2014:295-311.
- [32] NAJA-JAZZAR H,DE SILVA N,SKAF-MOLLI H,et al.OntoReSt:a RST-based ontology for enhancing documents content quality in collaborative writing[J].INFOCOMP Journal of Computer Science,2009,8(3):1-10.
- [33] CIANCARINI P,IORIO A D,NUZZOLESE A G,et al.Semantic annotation of scholarly documents and citations[C]//Conference of the Italian Association for Artificial Intelligence.[S.l.]:[s.n.],2013:336-347.
- [34] BARTALESI V,LOCURATOLO E,VERSIENTI L,et al.A preliminary study on the semantic representation of the notes to Dante Alighieri's Convivio[C]//Dh-Case 13 International Workshop on Collaborative Annotations in Shared Environment:Metadata.Vocabularies and Techniques in the Digital Humanities,2013:1-7.
- [35] BERTIN M,ATANASSOVA I.A study of lexical distribution in citation contexts through the IMRaD standard[C]//The Workshop on Bibliometric-Enhanced Information Retrieval Co-Located with European Conference on Information Retrieval.[S.l.]:[s.n.],2014.
- [36] SATELI B,WITTE R.Semantic representation of scientific literature: bringing claims, contributions and named entities onto the Linked Open Data cloud[J].PeerJ Computer Science,2015(1):e37.
- [37] MALHOTRA A,YOUNESI E,GURULINGAPPA H,et al. 'Hypothesis Finder:a strategy for the detection of speculative statements in scientific text[J].PLoS Computational Biology,2013,9(7):e1003117.
- [38] ATTWOOD T K,KELL D B,MCDERMOTT P,et al.Utopia documents: linking scholarly literature with research data[J].Bioinformatics, 2010,26(18):i568-i574.
- [39] CONSTANTIN A,PETTIFER S,VORONKOV A.PDFx:fully-automated PDF-to-XML conversion of scientific literature[C]//Proceedings of the 2013 ACM Symposium on Document Engineering.ACM,2013:177-180.
- [40] CASTRO L J G,MCLAUGHLIN C,GARCIA A.Biotea:RDFizing PubMed central in support for the paper as an interface to the web of data[J].Journal of Biomedical Semantics,2013,4(1):1-22.
- [41] PARINOV S,KOGALOVSKY M.Semantic linkages in research information systems as a new data source for scientometric studies[J]. Scientometrics,2014,98(2):927-943.

作者简介

王晓光, 男, 1978年生, 博士, 教授, 博士生导师, 研究方向: 数字信息资源管理、信息组织与知识服务、语义出版, E-mail: wxguang@whu.edu.cn。
宋宁远, 男, 1991年生, 博士研究生, 研究方向: 数字出版、信息组织与知识服务。

Review on the Scientific Paper Component Ontologies

WANG XiaoGuang^{1,2}, SONG NingYuan¹

(1. School of Information Management, Wuhan University, Wuhan 430072, China;
2. Center for Study of Information Resources, Wuhan University, Wuhan 430072, China)

Abstract: Scientific paper component ontologies standardize the definition of components and relationships within scientific articles. Reviewing and comparatively analyzing these ontologies is significant for the fully understanding of the multi-dimensional semantic attributions within scientific articles, facilitating the progress of semantic publishing systems. Based on the review about the development of scientific paper component ontologies, six typical ontologies that were put forward recently are chosen as research objects, comparatively analyzed the components and relations that these ontologies defined, explored the fundamental theories and the applications. There are varied types of scientific paper component ontologies, which are lack of an unified fundamental theories to support the standard definitions about the classes and the properties, restricts the popularization and application. In the future, the user demand of component ontologies should be taken into consideration, the direction of the development should be clarified to improve the usability of component ontologies.

Keywords: Semantic Publishing; Scientific Paper; Content Ontology; Paper Component; Comparative Analysis

(收稿日期: 2017-08-09)