

古籍注疏的知识表示与语义化建模研究

王晓光 翁梦娟 侯西龙 雷珏莹

摘 要 注疏是对古代典籍注释及再注释而形成的文本,不仅反映注疏者对古籍文本的认知理解,也是后人理解、传承与传播思想与文化的重要基础。利用本体和纳米出版物等语义技术对注疏文本进行知识表示和语义化建模,可以揭示注疏文献中蕴含的知识间的语义关系,并实现注疏文献的语义化出版与再造。为验证注疏知识表示和语义化建模方法的可行性及实用性,本文设计了包含引用关系的阐释本体,并以部分注疏文本为语料,实现了以纳米出版物为独立出版单位的注疏语义化表示与引用关系推断。实验证明,阐释本体可作为单语篇注疏知识单元结构化和跨语篇注疏知识单元关联化的数据模型,助力注疏文献的数据化处理与价值增值。注疏知识的语义化表示路径可以为古籍知识库建设、语义出版和数字化再造提供参考。图6。表5。参考文献41。

关键词 古籍注疏 阐释本体 知识表示 纳米出版物 古籍整理

分类号 G254

The Knowledge Representation and Semantic Modeling of Ancient Books Commentaries

WANG Xiaoguang, WENG Mengjuan, HOU Xilong & LEI Jueying

ABSTRACT

This paper proposes a semantic representation technical path of commentary knowledge based on ontology and nanopublication, aiming at realizing the minimal publication of commentary knowledge based on revealing the internal semantic relationship of ancient books commentaries and ensuring the traceability of the main responsible entities related to re-publishing activities and the credibility of publication contents. Five steps are included in the path: firstly, identify the minimized knowledge unit from the ancient books commentaries; secondly, realize attributes and resource extraction; thirdly, annotate the semantic relationship between resources and attributes; fourthly, fill the nanopublication template; and finally, generate credible URIs for nanopublications.

To illustrate the feasibility and practicability of this path, firstly, we identify four types of knowledge units from the ancient books commentaries: interpretation unit, citation unit, provenance unit, and alignment unit. Then, we construct the hermeneutics ontology to describe the knowledge units. The SWRL rules that infer author reference relations and ancient book reference relations from the sentence reference relations are included in the ontology. And then, we use the excerpts of some ancient books commentaries as a corpus to realize the semantic representation with nanopublication as an independent publishing unit and citation relationship inference. Finally, we use the MD5 algorithm to generate trusty URIs for nanopublication. Experiments have proved that the hermeneutics ontology can be a data model for the structuring of single-

通信作者:翁梦娟, Email: mengjuan_weng@whu.edu.cn, ORCID: 0000-0002-4631-5265 (Correspondence should be addressed to WENG Mengjuan, Email: mengjuan_weng@whu.edu.cn, ORCID: 0000-0002-4631-5265)

discourse ancient books commentaries and the association of cross-discourse ancient books commentaries, and that nanopublication can be a method for ensuring traceability and credibility.

The semantic representation technical path of commentary knowledge can provide a reference for the construction of smart data, semantic publishing and digital reconstruction of ancient books.

The innovation point of research lies in the fact that, for the first time, the nanopublication and the ontology are combined for the semantic representation of the ancient books commentaries. In addition, we also design the ontology for ancient books commentaries from the perspective of interpretation, which realizes the extensive semantic association between the ancient books commentaries and the external literature. 6 figs. 5 tabs. 41 refs.

KEY WORDS

Ancient books commentaries. Hermeneutics ontology. Knowledge representation. Nanopublication. Collation of ancient books.

0 引言

注疏是对前人所作典籍进行注释及再注释而形成的文本。不同时代的学者对典籍的解释对当时的社会产生着积极的意义,同时也反映着那个时代的价值理念^[1]。对今人来说,注疏是理解、传承与传播典籍中思想内涵的重要基础,也应当成为古文翻译和为古籍做新注的重要依据^[2]。

目前对注疏的访问、阅读与研究还是以纸本或数字化的文献为主,缺乏注疏文献与外部相关文献的关联、集成,为阅读理解与学术研究带来了诸多不便。具体表现为以下两个方面。一是同一典籍的注疏分散在不同的纸质或电子文献中,不利于针对同一典籍的全部注疏的汇聚。虽然集注或集释类文献能在一定程度上解决这一问题,但这类文献掺杂了作者的主观意图,部分注疏未被包含,而被后人忽略。二是纸质载体限制了注疏中的引用与被引文本间的即时访问,为文本的比对和参照阅读与引用计量分析带来障碍。古人在注疏时往往旁征博引,来提高注疏的可信性,这使得注疏中包含大量引经据典的内容。受限于纸质载体,注疏中对其他文本的引用无法以“跨越包含”的形式呈现,这既造成了注疏中的引用与被引文本存在内容差异,因为注疏中的引用不一定是对被引

文本的完全复制,也使得体现文化传承脉络的引用关系没有那么显而易见。

语义出版技术的发展为解决以上问题提供了可能性^[3],对注疏进行结构化再整理、多维化关联以支撑知识单元的汇聚重组有望成为现实。这不仅可以大大提高查找、阅读与整理注疏的效率^[4],还可以对经典注疏进行深度的引用分析,提高研究结论的可验证性^[3]。此外,在古籍数字文本的基础上,构建数据模型,并使用模型对古籍进行语义标注,以构建知识图谱或关联数据等知识库,进而支撑数字阅读与数字人文研究,也正在成为数字环境下古籍整理和数据化处理的新趋势^[3-9]。在此背景下,借助本体和关联数据等知识组织和语义建模技术实现对注疏知识的再整理,使之成为更有利于数字人文研究的智慧数据显得十分必要且可行^[10]。

纳米出版物模型的核心思想是通过知识单元及其语境信息的关联实现单个知识元的独立出版。纳米出版物是对纳米出版物模型进行形式化表示的新型出版物,这使得纳米出版物具有生成机器可读的关联数据的优势。从古籍整理的角度来看,对注疏内部语义的解构与再组织是一种数字化再造。纳米出版物为再造后产生的新的出版物责任主体的溯源、内容的可信保障提供了解决方案。

本文提出一种基于本体和纳米出版物的注疏文献数据化和知识语义化表示路径,并在着

重关注引用关系的基础上构建阐释本体,以期
为古籍的语义出版和智慧数据建设提供示范。
以《论语注疏·学而》《四书章句集注·论语·
学而》等文献为例,对《论语·学而》中“子曰:
‘学而时习之,不亦说乎?’”的注疏实现知识表
示和语义化建模,以验证路径的有效性。

1 研究综述

1.1 古籍注疏

注是对经书字句的注解,疏是对原文或注
的注解,注疏具有依附性。早期的注疏独自成
书,后为方便阅读,注疏逐渐附于原文之后;现
存注疏文献也多以“原文—注—疏”及其变形为
行文方式。随着时代发展,注疏的对象由经典
古籍扩大到文学作品,注疏体例也日益丰富,囊
括传、注、笺、疏、训、诂、考、证、音义、章句、解
诂、校注、义疏、疏证、说、论、衍义、评注、译注、
评议等^[11]。

注疏分为自注、他注、补注、集注四类^[12]。
注疏对象的粒度分为字、词、句子或句群、全篇。
注疏在理解原典中发挥着不同的作用,包括注
音、释义,阐释语法,考证和介绍作者生平、思
想、创作意图和书籍写作的背景,分析、评价和
发挥作品的思想意义,考据、说明、补充历史事
实和名物典故,文学艺术作品的赏析与评价,资
料补辑与辨析等^[11-13]。总的来说,注疏的作用
可分为三类:一是以训诂为主的解释,偏知识向
度;二是以义理发挥为主的阐释,偏意义向度;
三是训诂与义理发挥结合,同时具有知识向度
和意义向度。

注疏中通常包含大量对其他文本的引用。
从内容上看,引用内容包括对五经等原典的引
用、对他人注疏的引用以及对字典辞书等工具
书的引用。虽然在注疏中对其他文本的引用没
有严格的规范与标准,但从行文结构上看,引用
形式主要有两种:引用句的作者+引用句、引用
句原始来源古籍+引用句。

以上研究揭示了注疏的基本结构、体例、功

能以及注疏中引用的基本形式,为表征注疏知
识的本体构建提供了理论基础。由于现有研究
对注疏的体例种类及其发挥的作用还没有达成
共识,所以在构建本体时,需要保持本体模型的
可扩展性。

1.2 注疏的知识表示与语义建模

注疏具有规律的行文方式,可以看作是一
种半结构化文献。现有对注疏内容组织的研究
围绕着注疏语句展开,关注注疏的本体设计和
知识库建设。如马创新构建了训诂学本体^[14],
基于本体和 XML 实现对《论语集注》的半结构
化表示,将注疏文献中的被注疏字句与原典对
齐,借助注疏的位置相邻关系和内容等同关系
构建多种注疏文献之间的知识网络^[15,16];王
姗姗通过正则表达式匹配和相似度计算相结合
的方法实现《诗经》与《毛诗正义》的句子对
齐^[17];徐润华等基于相似度计算和回溯算法实
现《左传》与《春秋左传正义》的句子对齐^[18];周
好等使用 CRF、Bi-LSTM、SVM 和 BERT 四种模型
实现注疏文献中引书上下文的自动识别^[19];贾
凤旭将《周易》注疏文献的内部知识点分为注
疏对象、注疏、注疏目的和引用渊源,其中注
疏对象可以是字、词、短语或段落,注疏目的
包括文字辨析、词语明义、音读正讹、语法规
则、章旨文意和阐述典制名物等,引用渊源则
是记载原作者的注释语句及相关论述,可以辅
助典故、学说、研究成果等的源流考证^[20]。
在实践层面,中国哲学书电子化计划项目(cText)
在古籍数字文本的基础上,实现了注疏中引用
古籍名称与被引古籍间的超文本阅读^[21],但
这种关联仅限于文献层面,没有深入到引用句
层面。

现有研究着眼于注疏文献的半结构化行文
结构特征,设计训诂学本体,同时关注被注疏
字句与原典的对齐,这为同一典籍的注疏汇集
重组提供了可能,但忽略了两个重要内容:①没
有将注疏中的引用从注疏中分离出来,进而忽
略了对引用关系的建模,这使得单个注疏与其
他注疏或经典文献的关系无法得到揭示,也无
法

实现以引用为单元的分析;②从古籍整理的角度来看,对注疏文献内部知识单元的解构与再组织属于一种新的出版活动,是对注疏文献的再造,而再造后产生的新的出版物责任主体的可溯源性与内容的可信性没有得到保障。

1.3 纳米出版物

纳米出版物模型是21世纪初提出的一种全新的知识表示与语义模型,被广泛应用于生命科学、计算机科学、人文与社会科学等领域^[22,23]。“纳米”意指“具有科学意义的、机器可读的、最小的出版信息单元”。经过十余年的发展,纳米出版物模型已具备稳定的结构^[23-25],其核心思想是信息单元及其语境信息关联,通过信息单元及语境的形式化表示,支持对信息单元的整合、查询、推理等^[26,27]。具体来说,纳米出版物模型包括三个模块:论断模块、出处模块和出版物信息模块。每个模块均是一个RDF图,RDF图是“主—谓—宾”三元组的集合。其中,论断模块陈述纳米出版物声明;出处模块说明论断如何生成、由谁生成、何时生成、从何处抽取;出版物信息模块提供纳米出版物本身的出处信息,至少应包含创作者和时间戳^[26]。

根据Nanopublication.org网站发布的纳米出版物发布指南,格式良好的纳米出版物必须符合以下所有标准^[26]。①纳米出版物由一组RDF四边形组成(即主—谓—宾+语境)。②每个三元组的语境(如RDF图)必须指定为有效的URI。③有且只有一个“[N]rdf:type np:Nanopublication[H]”形式的四边形,其中[N]为纳米出版物URI,[H]为头部URI。④给定[N]和[H],有且仅有一个“[N]np:hasAssertion[A][H]”形式的四边形,其中[A]为论断URI;给定[N]和[H],有且仅有一个“[N]np:hasProvenance[P][H]”形式的四边形,其中[P]为出处URI;给定[N]和[H],有且仅有一个“[N]np:hasPublicationInfo[I][H]”形式的四边形,其中[I]为出版物信息URI。⑤[N]、[H]、[A]、[P]、[I]的URI必须都不同。⑥所有三元组必

须放在[H]或[A]或[P]或[I]之一。⑦[P]中的三元组至少有一个对[A]的引用。⑧[I]中的三元组至少有一个对[N]的引用。为便于检查纳米出版物是否被更改,强制要求纳米出版物的不变性,推荐使用可信URI^[28]。

可以看出,纳米出版物模型提供了对信息单元的组织框架,但没有对信息单元内的语义关系进行揭示。将纳米出版物模型与侧重于语义建模的本体结合使用可以克服这一不足^[23]。而且,纳米出版物的实现方式以及对可信URI的支持,可以实现对注疏文献再造后出版物责任主体的溯源,并保障出版物内容可信性。

借鉴以上研究成果,结合目前存在的不足,本文提出了基于本体和纳米出版物的注疏知识语义化处理方案。

2 基于本体和纳米出版物的注疏知识表示

纳米出版物在知识单元的最小化独立出版、关联数据生成、可溯源与可信性方面具有优势,本体在语义建模上具有优势,将两者结合并用于注疏知识的语义化表示,不仅可以在揭示注疏知识内部语义关系的基础上实现注疏知识最小化出版,还可以实现再出版活动相关责任主体的溯源、保障出版物内容的可信性。基于本体和纳米出版物的注疏知识语义化表示路径如图1所示。

语义化表示路径的输入包含三个部分。①本体。本体用于标注从非结构化的注疏文献库与经典古籍库中抽取的资源 and 属性的类别与关系。其中,领域本体用于建模注疏知识单元之间的语义关系;人物本体、书目本体及版本本体用于构建人物知识库和书目知识库;出版本体用于表征对注疏知识进行解构与再组织的相关活动。为表征注疏知识,设计了阐释本体作为领域本体。②注疏文献库与经典古籍库,这两个库是非结构化的古籍数字文本。③纳米出版物模板,该模板应符合格式良好的纳米出版物标准。

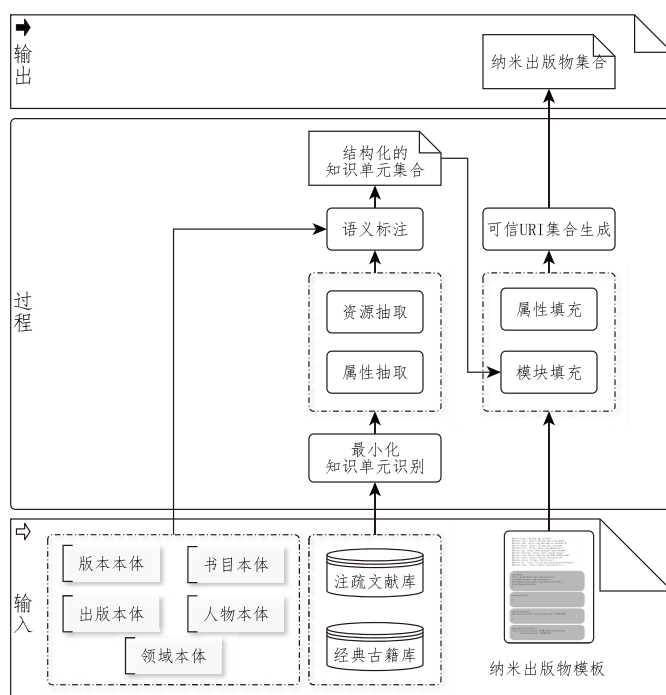


图1 基于本体和纳米出版物的语义化表示路径

语义化表示的最终结果是以纳米出版物为最小出版单位的知识库,具体过程可分为五步。

- ①从注疏文献库中识别最小化知识单元,旨在确定单个纳米出版物中应包含的陈述内容。
- ②资源与属性抽取,即从最小化知识单元中识别出可作为资源对象与属性的文本片段,并赋予资源对象 URI。
- ③语义标注,基于本体中的类和关系揭示资源对象的语义类别、资源对象之间以及资源对象与属性之间的语义关系。在标注之前,应进行标注实验或一致性检验,以说明本体的无歧义性。标注的结果是得到多个结构化的最小化知识单元。
- ④模板填充,包括模块填充和属性填充。首先判断结构化的最小化知识单元中单个三元组应置于哪个模块,并基于判断结果将其填充到相应模块;然后根据实际情况填充模板中的属性值,填充一个结构化的最小化知识单元后,得到一个纳米出版物,该纳米出版物包含 URI,但这一 URI 不能反映出纳米出版物内容是否更改,因此 URI 并不可信。
- ⑤生成可信 URI,借助哈希函数等签名加密技术

生成信息摘要,将该信息摘要附加到原 URI 上得到可信 URI。

3 古籍注疏的本体设计

本研究采用七步法^[29],以自上而下的方法构建表征注疏知识的本体。考虑到注疏是对原文进行阐释的结果,狭义的训诂学只涉及古代文献中的词义问题,广义的训诂学包括所有对古籍进行阐释的内容,而广义训诂学尚未形成系统的理论^[30]。如果用训诂学本体来组织注疏文献,可能会出现广义和狭义的混淆。因此本研究选择从阐释的视角^[31]理解注疏,在这一视角下,无论注疏以何种体例或何种类型出现,均可以将其视为对文本进行阐释的结果。本研究将设计的本体命名为阐释本体(Hermeneutics ontology,简称 HERU)。

3.1 阐释本体的设计要求

本研究在设计阐释本体时,着重关注注疏

中的引用关系。阐释本体设计的根本要求是最大化还原注疏文献中的结构语义,建立与外部文献的关联。根据注疏及其与外部文献的关联,将注疏的知识单元分为阐释单元、引用单元、出处单元、对齐单元四类。

(1) 阐释单元。究其本质,注疏的形成是文本阐释的结果。一旦作者完成文本创作,对文本意义的理解就不再是静态固定的^[32],作者意图也处于不确定的状态。由于知识结构和阐释方法的不同,针对同一文本,不同阐释者便会形成不同的注疏文本,由此注疏文本和被阐释的文本共同构成一个阐释单元。

(2) 引用单元。由于时间久远,“阐释者如何注疏”这一问题已不可能由阐释者本人回答。但是阐释者在注疏时有引经据典的习惯,这为人们从侧面回答这一问题提供了路径,即注疏中的引用。注疏文本及其中包含的引用文本共同构成一个引用单元。

(3) 出处单元。出处单元也包括两类:物理出处和原始来源。物理出处是指文本所在的物理实体文献,而原始来源则是文本首次形成时所在的文献。出处单元的意义在于描述文本的来源,为文本内容的准确性及可验证性提供保障。

(4) 对齐单元。对齐单元包括两类:被阐释

的文本与原典文本的对齐以及注疏中的引用文本与被引用文献中文本的对齐。前一类对齐单元的意义在于实现同一经典下相关注疏的自动汇集;后一类对齐单元的意义在于在注疏中的引用和原始文献之间建立关联,这样可以缩小印刷媒介下注疏中的引用与被引用文本之间的空间距离,进而使得引用与注疏文本的连接变得显而易见、易访问且可计算。

3.2 阐释本体的详细结构

3.2.1 核心概念及属性

为描述上文提到的四类知识单元,本研究设计的阐释本体包括四个一级类、四个二级类以及十个属性(见表1和表2),类与属性之间的关系如图2所示。阐释本体以直接复用本体中类或属性的方式复用了 FOAF、BIBFRAME、CITO^[33]、PROV、HiCO^[34,35]、OWL、Dublin Core 等本体与元数据标准,最终在 Protégé 5.5 中实现本体的形式化表示。

阐释本体表示四类知识单元的方式如下。

(1) 阐释单元。在本体中,“文本片段”一级类下定义了“阐释对象”和“阐释结果”子类。“阐释对象”和“阐释结果”均可以是任意长度的文本区域,以满足字、词、句、段、篇、标题等不同

表1 阐释本体中的类

一级类	一级类标签及说明	二级类	二级类标签及说明	复用本体
heru:TextSegment	文本片段:任意长度的文本区域	heru:OriginalText	原典文本:可独立存在的原创性作品,如《论语》《尚书》中的文本	—
		heru:InterpretationObject	阐释对象:被阐释(注或疏)的对象文本	—
		heru:InterpretationResult	阐释结果:对阐释对象进行阐释形成的结果文本	—
		heru:Citation	引用:阐释结果中引用的其他文献中的内容	—
foaf:Person	人:文本片段及其子类的创作者			FOAF
bf:Instance	实例:同一古籍的不同版本			BIBFRAME
bf:Item	单件:同一版本古籍的某一具体副本			BIBFRAME

表 2 阐释本体中的属性

属性	属性标签及说明	定义域	值域或值类型	复用本体
heru:hasInterpretation	有阐释	InterpretationObject	InterpretationResult	—
cito:cites	引用了	InterpretationResult	Citation	CiTO
prov:hadPrimarySource	原始来源是:文本首次出现时所在的文献	TextSegment 或其子类	Instance 或 Item	PROV
hico:isExtractedFrom	抽取自:文本当前的物理来源	TextSegment 或其子类	Instance 或 Item	HiCO
dcterms:creator	创作者:文本片段及其子类实例的创作者	TextSegment 或其子类	Person	Dublin Core
owl:sameAs	等同于:完全相同或基本相同	TextSegment 或其子类	TextSegment 或其子类	OWL
heru:typeOf	类型是:阐释对象的文本类型	InterpretationObject	xsd:string	—
heru:genreOf	体例是:古籍或阐释结果的体例	InterpretationResult 或 Instance 或 Item	xsd:string	—
heru:functionAs	功能是:阐释结果发挥的功能	InterpretationResult	xsd:string	—
prov:startedTime	发生时间:文本首次出版的时间	TextSegment 或其子类	xsd:gYear	PROV

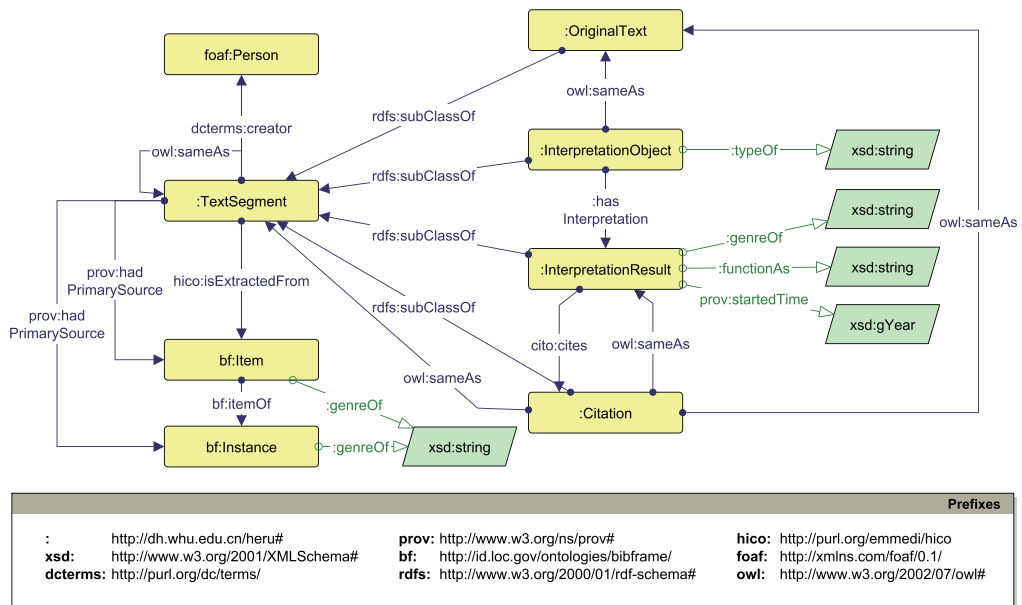


图 2 阐释本体的详细结构

类型的阐释单元表示。同时为了说明阐释对象属于哪种类型的文本,本研究在“阐释对象”类上定义了“类型是”属性。“类型是”用于说明对象可能表现为字、词、句、段、篇、标题等多种类型中的一种或多种。“阐释对象”和“阐释结果”之间的关系为“有阐释”。考虑到阐释结果具有章句、校注、传、笺、注、疏等不同的体例特征,而且不同的阐释结果在理解原典文本中发挥不同的功能,本研究在“阐释结果”类上定义了三个属性:“体例是”“功能是”和“产生时间”。“体例是”用于说明注疏的类型是传、注、笺、疏、训、诂、考、证、音义、章句、解诂、校注等多种类型中的一种或多种;“功能是”用于说明注疏发挥的作用可以是注音、释义,阐释语法,考证和介绍作者生平、思想、创作意图和书籍写作的背景,分析、评价和发挥作品的思想意义,考据、说明、补充历史事实和名物典故,文学艺术作品的赏析与评价,资料补辑与辨析等多种功能的一种或多种;“产生时间”是文本首次出版的时间,这与古籍的出版时间不同。“类型是”“体例是”“功能是”三个属性的值类型均为字符串;“产生时间”的值类型为数值。

(2) 引用单元。引用单元包含施引单元、被引单元及单元之间的关系。阐释本体中分别定义了“阐释结果”类和“引用”类来表示施引单元和被引单元,它们也是“文本片段”的子类。“阐释结果”类和“引用”类之间的关系为“引用了”。“引用了”这一对象属性的定义复用自 CITO。需要特别说明的是,从文本包含关系上来看,“引用”类的实例必然是“阐释结果”实例的子文本。举例来说,《论语注疏·学而》中有这样一个文本片段:“子曰:‘学而时习之,不亦说乎? 子曰:‘子者,男子之通称,谓孔子也。’王曰:‘时者,学者以时诵习之。诵习以时,学无废业,所以为说怍。’”其中“子曰……所以为说怍”是“阐释结果”类的实例,而“子曰……谓孔子也”“王曰……所以为说怍”均是“引用”的实例。这样的本体设计及标注有两个优点:一是通过简单查询即可确定阐释结果中是否包含引用;二是“引用”类的实例作为引用存在的依据,发挥着类似科学论文中参考文献的功能,便于考证引用的原

始来源。理想状态下,我们希望在“阐释结果”的实例与引用的原始文本间直接建立引用关系。但是现实情况中,先从阐释结果中分离出引用文本,再建立引用文本与其原始文本之间的等同关系,进而间接实现阐释结果与被引用文本的原始文本间的关联,更具有现实意义。

(3) 出处单元。为描述文本来源,本体中设计了三个出处属性:“创作者”“原始来源是”和“抽取自”。“创作者”是为了说明文本作者,后两种出处属性是为了区分不同来源的完全相同或大致相同的内容对象。比如“学而时习之,不亦说乎?”是从《论语》《四书章句集注·论语·学而》或《十三经注疏》等典籍中抽取到的,但其原始来源是《论语》。

(4) 对齐单元。对齐单元旨在建立不同来源但相似或相同内容文本之间的关联,关联关系用“等同于”表示。这种等同关系在阐释本体中表现为五种形式:①阐释对象与原典文本的等同;②引用与原典文本的等同;③引用与阐释结果的等同;④引用与文本片段的等同;⑤文本片段与文本片段的等同。其中阐释对象与原典文本之间等同关系的存在是因为现存注疏文献多为经、注、疏一体,这一等同关系可以实现原典与注疏文献的关联;引用与被引用文本之间等同关系的存在是因为部分注疏会引用《辞书》等工具书或五经等其他经典文献,这一等同关系可以实现被引文献与注疏文献之间的关联;引用与阐释结果之间等同关系的存在是因为部分注疏中会引用前人的注疏,这一等同关系可以实现注疏文献之间的关联,有助于梳理注疏文献的传承脉络;引用与文本片段的等同以及文本片段与文本片段的等同则是为兼顾其他可能出现的等同关系而设立的。

3.2.2 注疏文献引用网络的构建规则

从作者引用关系网络、古籍引用关系网络、引用句网络这三种引用网络出发开展研究,可以评估作者、古籍、引用句的重要性和影响力^[36,37]。本研究设计的阐释本体以引用关系为桥梁,结合阐释结果与引用的创作者和原始来

源,通过推理构建作者引用关系网络、引用句网络和古籍引用关系网络,支撑基于引用的计量研究。规则图式如图3所示,从左至右分别为作者引用关系、引用句关系与古籍引用关系。图中标签为 cito:cites 的实线为阐释本体中已有的

引用关系,标签为 cito:cites 的虚线为通过推理建立的引用关系,圆圈表示相应类的实例。

具体实现引用关系网络构建功能的 SWRL (Semantic Web Rule Language)推理规则见表3。表3与图3中的变量名称具有一致性。

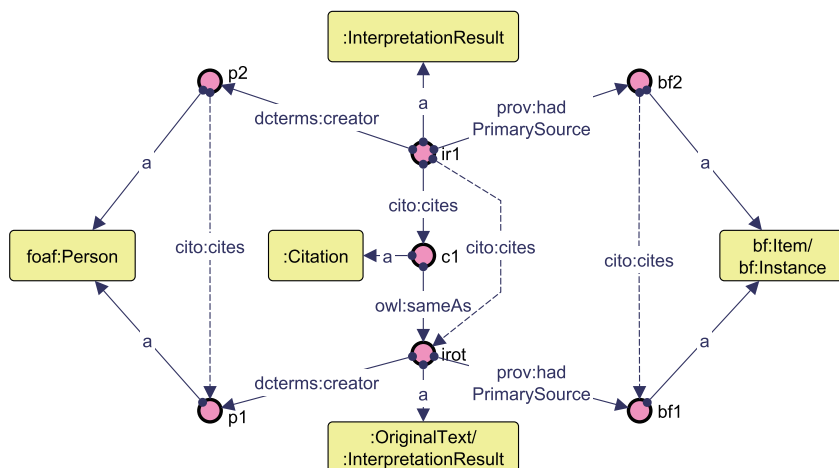


图3 引用网络构建的规则图式

表3 SWRL 推理规则

类别	SWRL 规则	说明
引用句网络	$InterpretationResult(? ir1) \wedge Citation(? c1) \wedge cito:cites(? ir1, ? c1) \wedge owl:sameAs(? c1, ? irot) \wedge (InterpretationResult \text{ or } OriginalText)(? irot) \Rightarrow cito:cites(? ir1, ? irot)$	$ir1$ 是“阐释结果”类的一个实例, $c1$ 是“引用”类的一个实例, $irot$ 是“阐释结果”或“原典文本”的实例。 根据 $ir1$ “引用了” $c1$, 而且 $c1$ “等同于” $irot$, 得出 $ir1$ “引用了” $irot$ 。
作者引用关系网络	$InterpretationResult(? ir1) \wedge cito:cites(? ir1, ? irot) \wedge (InterpretationResult \text{ or } OriginalText)(? irot) \wedge foaf:Person(? p1) \wedge dcterms:creator(? ir1, ? p1) \wedge foaf:Person(? p2) \wedge dcterms:creator(? irot, ? p2) \Rightarrow cito:cites(? p1, ? p2)$	为使阐述尽可能简洁, 这里直接基于已经得到的引用句网络进行说明, 且已出现的变量不再解释。 $p1$ 、 $p2$ 均是“创作者”的实例。 根据 $ir1$ 的“创作者是” $p1$, $irot$ 的“创作者是” $p2$, $ir1$ “引用了” $irot$, 得出 $p1$ “引用了” $p2$ 。
古籍引用关系网络	$InterpretationResult(? ir1) \wedge cito:cites(? ir1, ? irot) \wedge (InterpretationResult \text{ or } OriginalText)(? irot) \wedge (bf:Instance \text{ or } bf:Item)(? bf1) \wedge prov:hadPrimarySource(? ir1, ? bf1) \wedge (bf:Instance \text{ or } bf:Item)(? bf2) \wedge prov:hadPrimarySource(? irot, ? bf2) \Rightarrow cito:cites(? bf1, ? bf2)$	$bf1$ 、 $bf2$ 均是“实例”或“单件”的实例。 根据 $ir1$ 的“原始来源是” $bf1$, $irot$ 的“原始来源是” $bf2$, $ir1$ “引用了” $irot$, 得出 $bf1$ “引用了” $bf2$ 。

4 基于阐释本体和纳米出版物的注疏知识语义化表示

基于前文所述语义化表示路径,本研究以《论语注疏·学而》《四书章句集注·论语·学而》中对《论语·学而》中“子曰:‘学而时习之,

不亦说乎?’”的阐释为例,借助阐释本体实现了部分注疏文献的语义化表示。

4.1 实验数据准备

实验所使用的注疏文献和经典古籍文本数
据见表4。

表4 “子曰:‘学而时习之,不亦说乎?’”的部分语料

文 献	作者	年代	原 文
论语·学而	—	春秋	子曰:“学而时习之,不亦说乎?”
论语注疏·学而	何晏	三国	子曰:“学而时习之,不亦说乎?”马曰:“子者,男子之通称,谓孔子也。”王曰:“时者,学者以时诵习之。诵习以时,学无废业,所以说悦。”
	邢昺	北宋	
河南程氏经说·论语解	程颐	北宋	“学而时习之,不亦说乎?”习,重习也。时复思绎,浹洽于中,则说也。
河南程氏外书胡氏本拾遗	程颢	北宋	“学而时习之。”所以学者,将以行之也;时习之,则所学者在我,故说习如禽之习飞。
	程颐		
四书章句集注· 论语·学而	朱熹	南宋	子曰:“学而时习之,不亦说乎?”说,悦同。学之为言效也。人性皆善,而觉有先后,后觉者必效先觉之所为,乃可以明善而复其初也。习,鸟数飞也。学之不已,如鸟数飞也。说,喜意也。既学而又时时习之,则所学者熟,而中心喜说,其进自不能已矣。程子曰:“习,重习也。时复思绎,浹洽于中,则说也。”又曰:“学者,将以行之也。时习之,则所学者在我,故说。”谢氏曰:“时习者,无时而习。坐如尸,坐时习也;立如齐,立时习也。”(立:原作“一”。据清仿宋大字本改。)

实验所使用的领域本体为本研究所提出的阐释本体,使用的纳米出版物模板是基于格式良好的纳米出版物标准,在借鉴已有研究^[38]的基础上设计完成的。纳米出版物模板中包括命名空间、头部信息、论断、出处及出版物信息等模块,其中出处模块中至少包含论断由谁产生、出版物信息模块至少包含纳米出版物由谁在何时创建等内容。纳米出版物模板使用 Trig 语法(见图4)。

4.2 注疏知识语义化处理过程

(1) 识别最小化知识单元

考虑到现有注疏文献的形式,本研究根据单个纳米出版物所包含的陈述内容将最小化知识单元分为两类:①由单个阐释单元及相关的出处单元、引用单元组成;②单个对齐单元。

(2) 资源与属性抽取

从表4的数据中共识别出二十八个资源对象和两个属性,为资源对象赋予唯一标识。本研究为内容相同或相似但来源不同的资源赋予不同的 URI,这样做一是为了保留不同古籍中对同一内容的不同表达,方便参照对比;二是可以实现资源相关关系及属性的自动集成。

(3) 语义标注

标注一致性检验的具体过程为:本体设计人员根据图2向标注人员说明类和属性含义、介绍待标注文档,然后由三名标注人员独立完成标注。标注完成后,以本体设计人员的标注结果为标准,对标注人员的标注结果进行一致性检验。经计算,一致性程度分别为0.86、0.91、0.80,一致性程度均值为0.86,该值大于0.75,说明本体设计人员与标注人员对本体中概念与属性的理解具有较高一致性。



图4 用于注疏知识语义化表示的纳米出版物模板

不一致的内容主要集中在三种情形:“单件”与“实例”的混淆,等同关系的漏标,字或词作为阐释对象及相应阐释结果的漏标。与标注人员讨论后,总结出产生不一致的原因主要有三个。①未在标注前说明语料中文献名的区别,使标注人员混淆“单件”和“实例”。如果在标注之初便说明有的文献名有对应的实体,有的文献名仅是被某本书提及,没有对应的实体,则可以有效避免此类混淆。②不同标注人员对语料的阅读详细程度不一,导致部分等同关系以及字或词作为阐释对象及相应阐释结果的漏标。③等同关系是文本内容之间的等同,在本体中体现为类与类的等同,这使得部分标注人员对等同关系的概念产生了误解。虽然从本体设计角度来说,等同关系不应出现在本体中,因为等同关系体现的是类的实例与实例之间内容上的相似或相同关系,而不是类自身语义之间的相似或相同,但是为了说明引用数据构建的规则,本研究仍然在本体中保留了这一关系。

最后,基于本体设计人员的标注结果和三名标注人员的标注结果,确定最终的标注结果。

(4) 模块与属性填充及生成可信 URI

第一类最小化知识单元中的阐释单元和引用单元置于论断模块,出处单元置于出处模块;

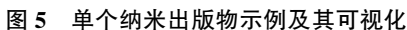
第二类最小化知识单元中的对齐单元置于论断模块。属性填充则根据实际情况确定。本研究使用 MD5 算法为纳米出版物生成信息摘要。

4.3 纳米出版物发布的结果

4.3.1 单个纳米出版物

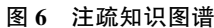
以朱熹在《四书章句集注·论语·学而》中对“说”的阐释“说,喜意也”“说,悦同”为例,说明第一类最小化知识单元的纳米出版物结构(见图5)。第二类最小化知识单元的纳米出版物结构与之类似,只是论断图中用到的类和属性及相关实例有所差别。本文中发布的纳米出版物均已通过 Nanopublication.org 的验证,符合纳米出版物发布规范。

图5的左上方为基于 Trig 语法的纳米出版物实现过程;右上方为该纳米出版物的可视化表达,它以人类可读的形式呈现纳米出版物的结构。从图5上半部分可以看出,发布的纳米出版物包括两个层面的知识图谱:①描述纳米出版物的创建依据、创作者和创建时间等元数据信息的纳米出版物图谱;②由论断子图和出处子图组成的、与注疏知识核心相关的注疏知识图谱。图5的下半部分是单个纳米出版物的注疏知识图谱。



点:①纳米出版物可以实现最小知识单元的独立发布与出版,可以通过修改纳米出版物实现对知识图谱内容的增、删、改;②纳米出版物的相关元数据可以用来支撑对论断可信度的评估和验证。

将表 4 中语料的标注结果全部发布为纳米出版物后,便得到以纳米出版物为最小独立出版单位的大规模注疏知识图谱(见图 6)。相较于直接创建知识图谱,其优势有两



可以用来支撑参照阅读和引用计量分析。最终得到的引用句数据、作者引用数据和古籍引用数据如表5所示。可以看出,这与表4呈现的基

本研究基于 SWRL 规则, 构建 SPARQL 查询式, 得到实验数据的引用数据。该引用数据

本事实是一致的,即数据中共存在五处引用:在引用句层面,三处引用没有找到对应的原始文本,有对应原始文本的两处引用中有一个与原始文本有些微差别;在作者引用层面,朱熹引用程子两次、引用谢氏一次,由于创作者数据的模糊,对马、王的引用缺少具体的施引对象;在古

籍引用层面,《四书章句集注·论语·学而》引用了《河南程氏经说·论语解》和《河南程氏外书胡氏本拾遗》,由于谢氏、马、王的内容缺失原始来源,导致部分引用关系缺少被引书籍,这些缺失数据提供了辑佚方向。

表 5 基于示例数据中的实例+规则得到的引用数据

层面	施 引	被 引
引用句	马曰:“子者,男子之通称,谓孔子也。”王曰:“时者,学者以时诵习之。诵习以时,学无废业,所以为说悖。”	—
		—
	说,悦同。学之为言效也。人性皆善,而觉有先后,后觉者必效先觉之所为,乃可以明善而复其初也。习,鸟数飞也。学之不已,如鸟数飞也。说,喜意也。既学而又时时习之,则所学者熟,而中心喜说,其进自不能已矣。程子曰:“习,重习也。时复思绎,浹洽于中,则说也。”又曰:“学者,将以行之也。时习之,则所学者在我,故说。”谢氏曰:“时习者,无时而不习。坐如尸,坐时习也;立如齐,立时习也。”(立:原作“一”。据清仿宋大字本改。)	习,重习也。时复思绎,浹洽于中,则说也。
		所以学者,将以行之也;时习之,则所学者在我,故说习如禽之习飞。
古籍引用	论语注疏·学而	—
	论语注疏·学而	—
	四书章句集注·论语·学而	河南程氏经说·论语解
	四书章句集注·论语·学而	河南程氏外书胡氏本拾遗
	四书章句集注·论语·学而	—
作者引用	何晏,邢昺	马
		王
	朱熹	程子
	朱熹	程子
	朱熹	谢氏

5 讨论

5.1 面向数字化再造的注疏语义出版平台建设

注疏之学是中国古典文献学的重要组成部分,注疏文献既是古籍整理工作的成果,也是古代学者在系统梳理与集成文献的基础上进行学术创新的集中体现。基于本文提出的注疏知识语义表示路径,对注疏文献进行知识解构与数

字化再造,古籍整理人员可以将个人对注疏知识内涵的理解转化为显性的语义关系,进而为更多研究人员提供结构化与智慧化的数据基础,也为使用人工智能技术进行文本分析与理解提供重要的语料。

古籍工作者通常对语义技术了解有限,注疏语义出版平台可以将复杂的技术流程转变为简单易用的操作界面,降低古籍整理人员的学习成本、提高古籍整理效率。目前,国内缺乏语义出版平台支撑。纳米出版物发布平台与现有

数字文本标注平台^[39]或知识图谱构建平台的核心区别在于对纳米出版物模板的支撑、可信 URI 的生成及版本变更。但是纳米出版物中命名图的使用导致无法利用现有的文本标注平台发布纳米出版物,如何在现有文本标注平台中增加相关功能以构建用户友好的纳米出版物发布平台,是未来应当关注的内容。

5.2 注疏知识单元的自动构建

本研究虽已通过部分示例呈现了注疏知识库及其应用的初貌,但数据量较为有限。未来对海量的注疏进行结构化、关联化、语义化表示仍面临诸多挑战。部分研究基于机器学习等技术实现了阐释对象与经典原文的对齐^[16-19],这些工作为大规模注疏文献的知识网络构建奠定了基础。

但在注疏知识单元的自动构建过程中,除了应关注阐释对象与经典原文的自动对齐外,还应关注阐释结果中的引文与被引用句的自动对齐、阐释对象及阐释结果的自动识别。cText 目前仅实现了引文与所在被引文献的超文本阅读,还未细化到引用句层面。这项工作难以推进的主要原因之一是同一典籍的不同注疏文献的数字化程度不一。目前,围绕同一典籍的不同注疏文献仍以纸质文献居多,这导致部分注疏文献内的阐释对象和阐释结果的自动识别工作无法推进,更不用说识别阐释结果中的引文了;而被引用的前人注疏所在的文献也未完成数字化,这导致阐释结果中的引文与被引用句的自动对齐缺乏数据基础。

5.3 注疏知识的交互式阅读

注疏知识的主要载体是注疏文献,目前注疏文献以印刷与数字等多种形态存在。纸质注疏文献仅支持线性阅读,而以数字化图像和文本为数据基础,实现超文本的交互式阅读,可以有效降低文献查找与集成时间、提高效率。

超文本的交互式阅读要求阅读系统的设计者为用户提供多种阅读路径以供选择。阐释本体中的语义关系为阅读路径的设置提供了更为

具体的方向,通过本体模型表达数据语义结构,使得人文学者可以根据需求实现资源的集成与重组(如注疏演化网络),进而支撑文本远读和辅助学术研究^[40,41]。人文学者更注重文本出处的原始来源,纳米出版物的组织框架则要求以形式化表达语境信息,这虽然不能消除语义标注过程中标注人员的主观性,但也使得责任主体及论断来源透明化,有助于人文学者做出更可信且可追溯、可验证的判断。基于超文本的交互式阅读还允许用户自由建立文本片段间的链接,以构建个性化的阅读路径。所以,注疏知识的交互式阅读还应关注如何让人文学者可以在文本细读过程中自由地实现语义化的关联、注释与评论,进而实现更具解释力、更可信的文本内容分析。

6 结语

本文提出基于本体和纳米出版物的注疏知识语义化表示路径,不仅可以在揭示注疏知识内部语义关系的基础上实现注疏知识最小化出版,还可以实现再出版活动相关责任主体的溯源,保障出版物内容的可信性。这一语义化表示路径可以为古籍知识库建设、语义出版和数字化再造提供参考。

此外,本文从阐释的视角理解注疏,并设计了用于表征注疏知识的阐释本体。这一本体通过阐释关系和引用关系将纷繁复杂的注疏文献与经典古籍关联起来,不仅可以实现不同粒度层面的跨语篇的关联与语义表示,还使得不同层面的引文计量分析成为可能。具体来说,阐释本体可以与实体识别、文本分类等自然语言处理技术结合,实现对注疏文献多种粒度、丰富关联关系的结构化表示,进而形成包括多种引用关系的大规模注疏知识网络。这种知识组织方式可以为注疏文献内多种对象的检索、浏览、汇聚与脉络梳理提供支撑,提高数字环境下古籍阅读、理解与为古籍做新注的效率。由于本研究设计的本体将注疏与被注疏对象的关系定义为更广泛的阐释关系,且注疏对象、注疏均可以是任意长度、

任意体裁的文本对象,而注疏对象的类型及体裁、注疏的体例及其发挥的功能则通过属性表征,所以该本体适用于各种类型的注释文献。但该本体仍存在一定局限性:本体中并未对体裁、功能及类型等属性的值域做出定义。

未来研究会针对本体中体裁、功能及类型

等属性的值域作出定义。此外,还将设计基于纳米出版物的文本语义标注与可视化工具,为基于纳米出版物的大规模注疏知识图谱发布提供基础设施支撑,并设计超文本的注疏交互式阅读平台,为人文学者和古籍爱好者阅读注疏文献提供便利。

致谢:本文系国家社会科学基金重大项目“文化遗产智慧数据资源建设与服务研究”(编号:21&ZD334)与湖北省自然科学基金创新群体项目“大数据资源语义组织与管理应用研究”(编号:2019CFA025)的研究成果。

参考文献

- [1] 陈少明. 经典注疏对现代人更有价值[EB/OL]. (2014-09-27)[2022-05-20]. <http://china.cankaoxiaoxi.com/2014/0927/511991.shtml>. (CHEN S M. Classic Zhushu are more valuable to modern people[EB/OL]. (2014-09-27)[2022-05-20]. <http://china.cankaoxiaoxi.com/2014/0927/511991.shtml>.)
- [2] 黄永年. 古籍整理概论[M]. 上海:上海书店出版社,2013:135-142. (HUANG Y N. An introduction to the arrangement of ancient books[M]. Shanghai:Shanghai Bookstore Publishing House,2013:135-142.)
- [3] 刘石,尹小林. 以数字映射古代文学经典[N/OL]. 光明日报,2022-03-23(11)[2022-04-21]. https://epaper.gmw.cn/gmrb/html/2022-03/23/nw.D110000gmrb_20220323_1-11.htm. (LIU S, YIN X L. Mapping ancient literature by digital technology[N/OL]. Guangming Daily,2022-03-23(11)[2022-04-21]. https://epaper.gmw.cn/gmrb/html/2022-03/23/nw.D110000gmrb_20220323_1-11.htm.)
- [4] 申利. 利用数字化资源提高古籍整理效率的实践和思考[J]. 图书情报知识,2012(5):120-125. (SHEN L. The practice and introspection of drawing upon digital resources to enhance the efficiency and quality of collating ancient books[J]. Document, Information & Knowledge,2012(5):120-125.)
- [5] 李明杰. 数字环境下古籍整理范式的传承与拓新[J]. 中国图书馆学报,2015,41(5):99-110. (LI M J. The inheritance and innovation of ancient book collation paradigm in the digital environment[J]. Journal of Library Science in China,2015,41(5):99-110.)
- [6] 陈力. 数字人文视域下的古籍数字化与古典知识库建设问题[J]. 中国图书馆学报,2022,48(2):36-46. (CHEN L. Digitalization of ancient books and classical knowledge repository from the perspective of digital humanities[J]. Journal of Library Science in China,2022,48(2):36-46.)
- [7] 刘石,李飞跃. 大数据技术与传统文献学的现代转型[J]. 中国社会科学,2021(2):63-81,205-206. (LIU S, LI F Y. Big data technology and the modern transformation of traditional philology[J]. Social Sciences in China,2021(2):63-81,205-206.)
- [8] 夏翠娟,林海青,刘炜. 面向循证实践的中文古籍数据模型研究与设计[J]. 中国图书馆学报,2017,43(6):16-34. (XIA C J, LIN H Q, LIU W. Designing a data model of Chinese ancient books for evidence-based practice[J]. Journal of Library Science in China,2017,43(6):16-34.)
- [9] 李明杰,张纤柯,陈梦石. 古籍数字化研究进展述评(2009-2019)[J]. 图书情报工作,2020,64(6):130-137. (LI M J, ZHANG X K, CHEN M S. Review on the research progress of the digitization of ancient Chinese books(2009-2019)[J]. Library and Information Service,2020,64(6):130-137.)
- [10] 曾蕾,王晓光,范炜. 图档博领域的智慧数据及其在数字人文研究中的角色[J]. 中国图书馆学报,2018,44(1):17-34. (ZENG L, WANG X G, FAN W. Smart data from libraries, archives and museums and its role in the digital humanity researches[J]. Journal of Library Science in China,2018,44(1):17-34.)
- [11] 董洪利. 古籍的阐释[M]. 沈阳:辽宁教育出版社,1997:83-231. (DONG H L. Interpretation of ancient

- books[M]. Shenyang: Liaoning Education Press, 1997: 83-231.)
- [12] 饶尚宽. 古籍语义阐释学[M]. 新疆: 新疆人民出版社, 1996: 11, 183-193. (RAO S K. Semantic hermeneutics of ancient books[M]. Xinjiang: Xinjiang Renming Publishing House, 1996: 11, 183-193.)
- [13] 张三夕. 中国古典文献学[M]. 武汉: 华中师范大学出版社, 2018: 255. (ZHANG S X. Chinese classical philology[M]. Wuhan: Central China Normal University Press, 2018: 255.)
- [14] 马创新, 陈小荷. 基于学科本体的训诂学知识组织体系初步构建[J]. 图书情报工作, 2013, 57(12): 118-122. (MA C X, CHEN X H. Primary construction of exegetics knowledge organization system based on discipline ontology[J]. Library and Information Service, 2013, 57(12): 118-122.)
- [15] 马创新, 陈小荷. 基于本体和XML的注疏文献的结构化知识表示[J]. 图书馆杂志, 2017, 36(8): 62-68. (MA C X, CHEN X H. Structural knowledge representation of commentary literatures based on ontology and XML[J]. Library Journal, 2017, 36(8): 62-68.)
- [16] 马创新, 陈小荷, 曲维光, 等. 《论语》与其注疏文献对齐语料库的构建[J]. 现代教育技术, 2012, 22(7): 109-113. (MA C X, CHEN X H, QU W G, et al. The construction of *The Analects of Confucius* and its commentary literatures alignment corpus[J]. Modern Educational Technology, 2012, 22(7): 109-113.)
- [17] 王姗姗. 《诗经》与其注疏文献的句子对齐研究[D]. 南京: 南京农业大学, 2018. (WANG S S. A study of sentence alignment in *The Book of Songs* and its commentaries[D]. Nanjing: Nanjing Agricultural University, 2018.)
- [18] 徐润华, 梁社会. 《左传》及其注疏文献的内容自动对齐研究[J]. 金陵科技学院学报(社会科学版), 2019, 33(2): 84-88. (XU R H, LIANG S H. Study on the automatic alignment of the contents between *Zuo Zhuan* and its annotated documents[J]. Journal of Jinling Institute of Technology(Social Science), 2019, 33(2): 84-88.)
- [19] 周好, 王东波, 黄水清. 古籍引书上下文自动识别研究——以注疏文献为例[J]. 情报理论与实践, 2021, 44(9): 169-175. (ZHOU H, WANG D B, HUANG S Q. Automatic recognition citation context in early Chinese literature; take the annotated literature as an example[J]. Information Studies: Theory & Application, 2021, 44(9): 169-175.)
- [20] 贾凤旭. 基于知识类聚的古籍知识库构建方法[J]. 图书馆学刊, 2015, 37(5): 45-48. (JIA F X. Construction method of knowledge base of ancient books based on knowledge clustering[J]. Journal of Library Science, 2015, 37(5): 45-48.)
- [21] 中国哲学书电子化计划. 论语注疏·学而[EB/OL]. [2022-03-03]. <https://ctext.org/lunyu-zhushu/xue-er/zhs>. (Chinese Text Project. *The Analects of Confucius—Xueer* [EB/OL]. [2022-03-03]. <https://ctext.org/lunyu-zhushu/xue-er/zhs>.)
- [22] 吴思竹, 李峰, 张智雄. 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, 39(4): 102-109. (WU S Z, LI F, ZHANG Z X. Research on semantic representation and publishing schema of knowledge resource; take nanopublication as an example[J]. Journal of Library Science in China, 2013, 39(4): 102-109.)
- [23] 王晓光, 宋宁远. 语义出版物的内容组织架构研究——基于纳米出版物和微型出版物的比较分析[J]. 出版科学, 2017, 25(4): 20-27. (WANG X G, SONG N Y. Content schema of semantic publication: a comparative analysis based on nanopublication and micropublication[J]. Publishing Journal, 2017, 25(4): 20-27.)
- [24] GROTH P, GIBSON A, STICKLER P. The anatomy of a nanopublication[J]. Information Services and Use, 2010, 30(1): 51-56.
- [25] MONS B, VELTEROP J. Nano-publication in the e-science era[C]//Proceedings of the Workshop on Semantic Web Applications in Scientific Discourse(SWASD 2009). Washington DC, USA.
- [26] GRAY A, BURGER K, KOTOULOS S, et al. Nanopublication guidelines[EB/OL]. (2023-05-03) [2023-05-16]. https://nanopub.net/guidelines/working_draft/.
- [27] 李春秋, 徐曾旭林, 宋宁远, 等. 基于纳米出版物的中文学位论文语义组织研究[J]. 中国图书馆学报, 2021, 47(5): 97-115. (LI C Q, XU Z X L, SONG N Y, et al. Nanopublication-based semantic organization of Chinese dissertation[J]. Journal of Library Science in China, 2021, 47(5): 97-115.)

- [28] TOBIAS K, MICHEL D. Trusty URIs: verifiable, immutable, and permanent digital artifacts for linked data[C]// Proceedings of the 11th European Semantic Web Conference (ESWC). Crete, Greece: Springer, 2014.
- [29] USCHOLD M, GRUNINGER M. Ontologies: principles, methods and applications[J]. The Knowledge Engineering Review, 1996, 11(2): 93-136.
- [30] 董洪利. 古典文献学基础[M]. 北京: 北京大学出版社, 2020: 270-274. (DONG H L. Fundamentals of classical philology[M]. Beijing: Peking University Press, 2020: 270-274.)
- [31] 张江. 构建当代形态的中国阐释学[N]. 中国社会科学报, 2021-10-18(006). (ZHANG J. Constructing contemporary Chinese hermeneutics[N]. Chinese Journal of Social Sciences, 2021-10-18(006).)
- [32] MASSER I, WEGENER M. What new heritage for which new Europe?[M]// ASHWORTH G J, LARKHAM P J. Building a new heritage: tourism, culture and identity in the new Europe. London: Routledge, 1994: 31.
- [33] PERONI S, SHOTTON D. FaBiO and CiTO: ontologies for describing bibliographic resources and citations[J]. Journal of Web Semantics, 2012, 17: 33-43.
- [34] DAQUINO M, PASQUAL V, TOMASI F. Knowledge representation of digital hermeneutics of archival and literary sources[J]. Italian Journal of Library, Archives, and Information Science, 2020, 11(3): 59-76.
- [35] DAQUINO M, TOMASI F. Historical context ontology (HiCO): a conceptual model for describing context information of cultural heritage objects[C]// 9th Research Conference on Metadata and Semantics Research. Manchester, UK: Springer, 2015: 424-436.
- [36] 黄水清, 周好, 彭秋茹, 等. 引书的自动识别及文献计量学分析[J]. 情报学报, 2021, 40(12): 1325-1337. (HUANG S Q, ZHOU H, PENG Q R, et al. Automatic recognition and bibliometric analysis of cited books[J]. Journal of the China Society for Scientific and Technical Information, 2021, 40(12): 1325-1337.)
- [37] 舒非, 丰鹏萱, 邱均平, 等. 基于我国古籍引经据典现象的引文分析研究[J]. 情报学报, 2021, 40(12): 1338-1346. (SHU F, FENG L X, QIU J P, et al. Exploring the function of citation using ancient Chinese literature[J]. Journal of the China Society for Scientific and Technical Information, 2021, 40(12): 1338-1346.)
- [38] TOBIAS K. Nanopublications representing the nanopublication template ontology (Nanopub Index) [EB/OL]. (2021-01-04) [2022-03-23]. <http://server.nanopubs.lod.labs.vu.nl/RAsWN6p7SfCDHxZRjaiajJdS76sd-zlEaT8sWKTUYK7z90>.
- [39] 魏希德. 中文与韩文数据集的创建、关联与分析——用 MARKUS 与 COMPARATIVUS 进行数字文本标注[J]. 数字人文, 2021(3): 28-36. (WEERDT H D. Creating, linking, and analyzing Chinese and Korean datasets: digital text annotation in MARKUS and COMPARATIVUS[J]. Digital Humanities, 2021(3): 28-36.)
- [40] 张琪, 王东波, 黄水清, 等. 史书多维知识重组与可视化研究: 以《史记》为对象[J]. 情报学报, 2022, 41(2): 118-129. (ZHANG Q, WANG D B, HUANG S Q, et al. Multi-dimensional knowledge reorganization and visualization of history books: based on *Records of the Grand Historian*[J]. Journal of the China Society for Scientific and Technical Information, 2022, 41(2): 118-129.)
- [41] 欧阳剑, 任树怀. 数字人文研究中的古籍文本阅读可视化[J]. 图书馆杂志, 2021, 40(4): 82-89, 99. (OUYANG J, REN S H. Visualization of ancient texts reading in digital humanities research[J]. Library Journal, 2021, 40(4): 82-89, 99.)

王晓光 武汉大学信息管理学院教授, 博士生导师, 武汉大学文化遗产智能计算实验室主任。湖北 武汉 430072。

翁梦娟 武汉大学信息管理学院博士研究生。湖北 武汉 430072。

侯西龙 曲阜师范大学传媒学院副教授, 武汉大学文化遗产智能计算实验室兼职研究员。山东 日照 276800。

雷珏莹 武汉大学信息管理学院博士研究生。湖北 武汉 430072。

(收稿日期: 2022-06-23; 修回日期: 2022-08-24)