

历史地图知识组织:需求、框架及实践

齐小英 杨海平

摘 要 历史地图是信息的重要载体之一,其信息隐含在由多种元素组成的复杂网络中。为了高效发挥历史地图的地理空间表示作用,深入挖掘其历史价值、人文价值与证据价值,需要构建统一的历史地图知识组织框架与语义组织模型。本文从形态描述需求、内容描述需求、地理知识服务、文化知识服务、证据知识服务五个维度对历史地图的服务目标进行系统分析,获得“读图者”与“资源管理者”对历史地图的知识组织需求。在此基础上,设计了包含历史地图文本化、概念识别、本体构建以及知识关联路径与关联模式等核心环节的历史地图知识组织框架,并阐述各环节及其之间的核心含义与逻辑关系,形成了历史地图的结构化、数据化、语义化组织机制。本研究为历史地图提供了完整的语义描述范式,丰富了多源异构信息载体的知识组织理论与方法。图3。表5。参考文献41。

关键词 历史地图 知识组织 知识关联 本体 语义组织

分类号 G254 K992.6

Historical Maps Knowledge Organization: Need, Framework and Practice

QI Xiaoying, YANG Haiping

ABSTRACT

Historical maps are one of the important carriers of information, and their information is implicit in a complex network composed of multiple elements. In order to effectively play the geographical spatial representation role of historical maps and deeply explore their historical, humanistic and evidential values, it is necessary to propose a unified historical maps knowledge organization framework and semantic organization model.

This paper systematically analyzes the service goals of historical maps from five dimensions: morphological description needs, content description needs, geographical knowledge services, cultural knowledge services, and evidence knowledge services, and obtains the understanding of historical maps by “map readers” and “resource managers”. On this basis, a historical maps knowledge organization framework was designed that includes core links such as historical maps textualization, concept identification, ontology construction, and knowledge association paths and patterns. It also elaborated on each link and the core meanings and logical relationships between them. A structured, data-based, and semantic organization mechanism for historical maps has been formed.

The research provides a complete semantic description paradigm for historical maps and enriches the knowledge organization theory and methodology for multi-source heterogeneous information carriers. In the future, it is still necessary to strengthen the automatic generation method of historical maps structured text based on deep learning to improve the efficiency of textualization. Meanwhile, we should expand the association from traditional resources such as documents and images to new resources such as oral history, social medias, knowledge maps, VR to

通信作者:杨海平,Email:yhp@nju.edu.cn,ORCID:0009-0009-4027-5950(Correspondence should be addressed to YANG Haiping, Email:yhp@nju.edu.cn,ORCID:0009-0009-4027-5950)

achieve the multi-modal knowledge association of historical maps. 3 figs. 5 tabs. 41 refs.

KEY WORDS

Historical maps. Knowledge organization. Knowledge association. Ontology. Semantic organization.

0 引言

2022年4月,中共中央办公厅、国务院办公厅印发《关于推进新时代古籍工作的意见》,要求档案、方志、古地图等工作主管部门加强本领域古籍工作,积极开展古籍文本结构化、知识体系化的研究和实践,统筹古籍数字化版本资源建设与服务,推进古籍专业数据库和中国古籍版本目录知识系统建设,着力构建古籍知识服务体系。在历史地理学科领域,历史地图指“今人”以其所在时代的当代地图为底图,依照“今人”的需要,以历史上某一年代或时期的地理状况为内容,以地图形式呈现的学术性成果^[1]。本文的“历史地图”是宏观层面上历史时期所产生的地图统称,包括古地图、舆图、其他类型地图以及上述所指历史地图等,是古籍资源的重要组成部分。历史地图是记录时空分布的载体,也是珍贵的历史文化遗产与重要的领土划分证据。对历史地图进行知识组织不但能够高效地发挥地图地理空间表示的作用,更能深入挖掘其历史价值、人文价值与证据价值,提供多元的知识服务。

在大数据与新文科背景下,知识服务对数据组织有特定的要求,不仅需要通过数据间的关联形成知识,更需要在知识服务的框架下构成知识组织体系^[2]。因此,发挥历史地图资源价值的关键在于历史地图的数字化、数据化以及知识化,其核心环节是构建历史地图的知识组织框架与语义组织模型。目前,各行业和各领域根据自身需求已经发布了很多知识组织与服务的标准^[3-4],且随着用户对知识服务的需求不断增加、领域知识组织方法不断向更高层次发展,知识组织的对象也从文献^[5]逐渐拓展至视频^[6]、图像^[7-8]等各类信息载体。部分学者针对历史地图开展了基于元数据、文本注释、图像识别的地图知识挖掘^[9-12]以及基于本体和语义规则的地图语义分析^[13-15],但由于用户知识需求的内容和需求表达的详略程度各异、研究路径不一等问题,目前尚未形成系统且满足各方需求的历史地图知识组织理论和方法。与用文字直观表示信息的历史文献和档案不同,历史地图所表达的信息隐含在线画、影像、符号、数字及其关系之中,并与地图来源信息、图例、比例尺、投影方法、坐标、注记文字等形成复杂的信息网络,需要经过读图者提取与整理才能获得有效信息。现有数据库对于历史地图的索引与描述规则主要针对著录资料来源及包括文献细节项、载体形态描述项等在内的书目信息,对于历史地图深层次内容语义的表示能力不足。同时,由于历史地图信息的特殊性与隐含性,现有针对历史文献与档案的知识组织模型无法直接应用在历史地图知识组织。除了信息的挖掘与表示以外,语义信息的共享以及智能重用也是历史地图知识组织的主要挑战之一,建立地图与地图、地图与其他信息载体之间的知识关联是新文科背景下知识组织的发展趋势,故需要构建一套统一、规范的历史地图知识组织框架,实现历史地图资源的序化与关联。

为了推动历史地图资源数据化与关联化,支撑历史地图的知识服务,本文围绕历史地图知识组织的用户需求,提出历史地图的知识组织框架,阐述知识组织的实践与服务路径,形成历史地图的结构化、数据化、语义化机制,丰富多源异构信息载体的知识组织理论与方法。

1 历史地图知识组织的需求阐释

知识组织理论与方法服务于特定的应用目标,历史地图知识组织也需要立足于用户需求。在

《图书馆·情报与文献学名词》一书中,“用户需求”的定义为:“用户对各种信息及相关服务的要求的总称,充分了解用户的信息需求是做好图书情报服务工作的前提。”不同的需求主体(用户),对于历史地图知识组织的需求层次和目的不同。“读图者”往往关注历史地图内容的结果呈现,强调内容描述的细粒度、多层次、多角度、精准化;“资源管理者”主要关注历史地图的资源建设、内容揭示和全面整合,构建满足“读图者”需求的知识内容和资源体系。但在实际工作中,“读图者”和“资源管理者”并没有非常明确的界限,也存在“读图者”本身就是“资源管理者”的现象,二者关于历史地图知识组织的需求较统一。因此,本文重点强调“读图者”与“资源管理者”关于历史地图知识组织的知识描述需求与知识服务需求,而不对二者需求进行详细区分。

1.1 知识描述需求

以往建立历史地图数据库的目的在于对地图进行保存、分类与检索。在大数据时代,随着“读图者”需求的不确定性、动态性与个性化,“资源管理者”不但要为“读图者”提供关于该历史地图“来自哪里”“由谁绘制”“如何绘制”等浅层形态问题的解答,更要回答该历史地图“有什么”“说明什么”等深层内容问题。由此可见,用户对历史地图知识组织的知识描述需求应包括形态描述需求与内容描述需求。

1.1.1 形态描述需求

表1总结了现有数据库与元数据描述规范中关于历史地图的元数据描述项。尽管不同数据库和描述规范的元数据核心元素及数量略有差别,但均集中于历史地图的来源、收藏和馆藏信息(包括题跋印记、获得方式、馆藏号等)和历史地图的客观属性(名称、出版机构、类型、性质、色彩、图幅等)。本文将上述信息统称为形态描述。形态描述需求是对历史地图的显性形态化知识的必要描述,往往与历史地图的真实性与可靠性密切相关,是用户认识历史地图价值的初始标准。因此,历史地图知识组织必然要对地图的形态描述需求进行知识表示,以此作为历史地图检索、分类与关联的数据基础。

表1 历史地图元数据描述项

数据库及元数据描述规范	描述项
美国国会图书馆历史地图数据库	名称、描述、贡献者、出版时间、主题类型、分类号、馆藏地、馆藏号、格式等
大卫·拉姆齐历史地图数据库	名称、出版信息、类型、尺寸、比例尺、国家、主题、分类号、描述等
得克萨斯大学图书馆历史地图数据库	名称、出版者、创建时间、格式、类型、主题、来源、标识码等
英国皇家地理学会历史地图数据库	编码、名称、出版时间、馆藏地、馆藏号、保存格式、状态、评论等
DC 元数据	日期、创建者、主题、出版者、类型、描述、其他责任者、格式、来源、权限、标识符、语种、关联、覆盖范围等
VRA 元数据	格式、方法、材质、地域、年代、主题、创作者、关系、版权构成等
舆图元数据	题名、责任者、日期、出版者、附注、相关资源、主题、时空范围、语种、类型、标识符、馆藏、收藏、版本、载体形态、制图、权限等
中国国家图书馆:《舆图要录》	图名、著者、版本源流、绘制及出版年代等
澳门科技大学:澳门古历史地图	编码、名称、来源、出版时间、描述、贡献者、类型、主题词、语种、格式、比例尺等
中国台湾中研院:历史地图数位典藏	名称、性质、类型、贡献者、出版时间、保存状况、处理状况、管理项目等

本文以《国家图书馆舆图元数据规范与著录规则》(2014)所提出的舆图资源内容和形式特征的元数据标准^[16]为基础,综合表1中的元数据描述项,总结出形态描述需求的具体内容包括基本属性、收藏属性、文物属性、制图属性以及出版属性,具体细分项如表2所示。相比表1中的描述项,本文所提出的形态信息增加了来源(私人收藏、个人购买、网络下载)、保存(文物级别、破损级别)、参数(画幅、色彩、尺寸、投影方法、地形表示法、坐标系)以及数量(幅、册、种、件、本)等描述项,进一步反映历史地图的制作背景、制作技术以及文物价值等。

表2 历史地图形态描述需求细分项

内 容	元 素	元素修饰词	内 容	元 素	元素修饰词
基本属性	名称(title)	图名	收藏属性	馆藏(collection)	馆藏单位
		译名			馆藏地址
	语言(language)	语种			馆藏URI
					馆藏号
	形式(form)	单图	文物属性	保存(protection)	文物级别
		插图			破损级别
		挂图			
	类型(type)	地图集	制图属性	参数(parameter)	画幅
		行政区划图			色彩
		疆域变迁图			尺寸
		自然资源图			投影方法
		地形图			地形表示法
		航海图			坐标系
			出版属性	机构(agency)	出版者
	日期(date)	绘制日期(年号)			出版地
		绘制日期(公元)		记录(record)	卷号
		出版日期(年号)			页码
		出版日期(公元)			版本
		修订日期(年号)		数量(quantity)	幅
		修订日期(公元)			册
	收藏属性	私人收藏			种
		个人购买			件
		网络下载			本

1.1.2 内容描述需求

计算机技术和地理信息技术的应用使一系列的信息转化行为变得容易,但应尽量减少核心信息的损失^[17]。因此,历史地图的内容描述应围绕其核心内容展开。地理概念是认知空间的要素,地图符号语义即这些符号所代表的实际地理概念或空间特征的含义,地理概念和地图符号语义是相互依存的,地理概念为地图符号提供了语义基础,而地图符号则通过其语义内容来传递地理概念,这种相互作用构成了地图作为地理信息传递工具的核心^[18-19]。因此,历史地图的知识组织应着重关注深层次的语义知识,即内容描述需求,以弥补形态描述信息在表示历史地图的绘制背景、绘制立场、绘制目的、绘制价值等内容上的不足。内容描述要求知识组织者对历史地图抽象符号表达的隐性信息进行

挖掘,显式表示成具有明确语义的知识元。历史地图的内容描述需求可以细分为空间知识、图例知识、颜色知识、线画知识、注记知识、附图知识与分类知识,具体阐述如下。

(1) 空间知识

历史地图的基本功能是记录并描述由点、线、面所表达的空间属性与关系,空间知识是历史地图内容描述的基本内容。空间知识具体可以分为点(地点)、线(界线)、面(空间关系)。其中,地点指历史地图上所标注的各种地理实体对象;界线指历史地图上所绘制的各种区域、海域所涉及的国际线、省界线、行程线、航线、洋流线、风向线以及贸易线等;空间关系指历史地图中所表示的各种拓扑关系(如在……之内、在……之外等)、方位关系(如东北、东南、西北、西南等)与度量关系(如距离)。地点、界线与空间关系知识不但可以用于常规地理知识服务,更能构成历史人文演变与领土主权历史归属分析的空间基础,以作为文化知识服务与证据知识服务的空间知识要素。

(2) 图例、颜色与线画知识

图例知识标示历史地图各种抽象符号和元素的意义,是“全图之精神”^[20],各种线段的含义往往在图例中标示出来。颜色知识指图中文字、地点、区域以及海域的绘制颜色,例如,采用红色注明某地名;采用中国本土米黄色标绘某岛屿颜色。线画知识指历史地图中绘制的各种场景或图案(如正在航行中的帆船),在地图中具有一定的语义解释作用。

(3) 注记知识

注记既包括历史地图上体现对象的名称、性质、数量等特征的“名称”注记,也包括历史地图上体现与对象相关的背景、事件、价值等特征的“说明”注记。历史地图上的注记知识可以看作地图图例的另一种形式,是地图符号的补充,往往反映一般符号难以传递的信息。例如,《太平洋现势图》在海域中标明了“危险地带”“破船残骸”“码头遗址”等字样,直观表达了海洋地理、遗址分布等信息。此外,与历史地图本身相关的文献、条约等文字,也可看作历史地图的附加说明注记。

(4) 附图知识

附图指历史地图上绘制的“特定范围内的地图”,一般位于地图的左下角或右下角。附图包含了比主图分辨率更高、粒度更细的各种知识,往往能够清晰地表达主图无法表达的地理关系。

(5) 分类知识

分类知识指基于权威的分类法体系——《中国图书馆分类法》对历史地图进行分类,主要表现为一组分类代码。由于每一张历史地图可从不同角度解读,故分类号可以有多个。例如,1905年出版的《印度洋自然地图》,既属于K993(各国地图),也属于P98(自然地理图),还属于P724(印度洋)。历史地图分类知识便于用户快速掌握历史地图的性质与主题,进一步实现语义分类。

1.2 知识服务需求

国内外有关历史地图的知识服务应用广泛,总体包括地理知识服务、文化知识服务、证据知识服务。这三类知识服务是我国历史大数据资源建设的重要目标和文化产业数字化转型面临的重要环节,是当代新文科建设的学术前沿^[21]。

(1) 地理知识服务

历史地图的地理知识服务指通过建立核心地理要素之间的知识关联,表达复杂的时空关系并实现自动推理^[22-23],以支持地理位置查询与地理特征分析等服务。由于历史地图元素是空间对象的抽象符号,具有相对独立性与语义不确定性,计算机无法直接识别由地图元素相互组合所表达的深层知

识。因此,如何将历史地图语义知识转换为计算机能够识别的结构化信息,是解决地理信息语义不确定性、语义共享和互操作问题的关键^[24-25]。例如,基于地理信息技术实现古今地图配准与映射,构建统一的地理定位坐标系统与时空关联模型,实现对某幅未标注经纬度的历史地图中的地点进行定位,支撑特定地点古今位置空间对比分析与地名演变研究。

(2) 文化知识服务

历史地图的文化知识服务指通过挖掘历史地图蕴含的人文历史信息,溯源考证不同历史时期的地名、社会经济变迁等,从而发挥历史地图的文化遗产价值。例如,《宋代南海交通图》标绘了菲律宾、占城、三打根、渤泥、婆罗洲、真腊等国,注记包括玳瑁、丁香、胡椒、檀香、象牙、麝香、犀角等物品,反映了我国宋代海上丝绸之路的商业文化。又如,《郑和航海图》包括 20 页航海地图、109 条针路航线、4 幅过洋牵星图,体现了当时先进的航海技术和海洋地理科学水平。除此之外,也有如《西南洋各番针路方向图》与《东洋南洋海道图》等海道图,以黄色圆点组成弦月状(注记“气”)和长条状的岛屿(注记“长沙”“石塘”)^[26],这种山水画式地图,色调和谐、层次有别,蕴含极大的文学艺术价值和明朝海上文学文化分析价值^[27]。

(3) 证据知识服务

每一幅历史地图都是国家版图的重要表现形式^[28],是国家领土主权归属的有力补充证明。众多历史事实证明历史地图在维护国家领土主权方面发挥着重要的证据作用^[29]。历史地图的证据知识服务指将地图知识组织成证明特定历史事实的证据链条,以构成划定领土归属的历史和法理基础,用于佐证地域归属、解释边界条约等。影响历史地图证据效力的五个因素包括地图是否清楚地反映国家主权意图、地图制作是否精确、地图本身是否足够中立客观、一国官方地图是否前后一致、一国地图是否得到国际社会以及利害关系国家的承认或默许等^[30]。在国际法院判决的领土及边界争端实际案件中,几乎都涉及大量的历史地图证据^[31],例如 1928 年美国/荷兰的“帕尔马斯岛主权争端案”、1986 年“布基纳法索/马里边界纠纷案”、1999 年“卡西基利/塞杜杜岛案”等。

综上所述,地理知识服务、文化知识服务以及证据知识服务不但要求充分挖掘和分析历史地图包含的形态知识与内容知识,还需要将多幅历史地图知识关联成知识网络,以满足“读图者”与“资源管理者”多样化和个性化的历史地图知识需求。

2 历史地图的知识组织框架

知识组织以知识的有效转化与服务为目标,历史地图的知识服务应与“资源管理者”和“读图者”的知识描述需求和知识服务需求相匹配。因此,整个历史地图知识组织框架紧密围绕用户知识描述需求和知识服务需求进行设计。本文提出的历史地图知识组织框架如图 1 所示,此框架围绕“资源管理者”和“读图者”两类用户的知识描述需求和知识服务需求,以知识表示和知识关联两个阶段为指导路径。

知识表示是将单幅历史地图的抽象信息转换为具有明确语义的知识单元,并抽取核心概念与语义关系的过程,包括历史地图文本化、概念识别以及本体构建等环节。历史地图的信息隐藏在抽象符号之中,为了有效地利用历史地图信息,地图的语义文本成为恢复任何语义歧义和构建历史与当前地理空间之间关系的必要条件^[32]。在进行概念识别之前,必须将难以分析的符号转换为可计算、可分析、具有明确语义的文本,即历史地图文本化过程。整个历史地图知识表示过程如下:首先从抽象的历史地图符号中提取出具有明确且最小单位的知识元,将其显式表示成结构化文本之后,进行句子级别的知识单元解析;之后,再从句子级别的知识单元中识别语义角色并抽取语义依存关系,进行概

念识别;最后,采用自顶向下的原则进行概念整合或分解,建立概念及其之间的关系,构建历史地图本体模型。

知识关联指在创造和利用知识的活动中,知识因某种内在或外在的联系而显示关联的行为及状态,是语义信息上的多角度、多层次关联^[33]。历史地图的知识关联强调由单幅或多幅地图共同呈现综合知识,将历史地图中蕴含的语义重组成语义更丰富、逻辑层级更深、价值更高的知识网络,以满足知识服务需求,具体包括内部关联与外部关联两种关联路径以及基于概念与基于实例两种关联模式。

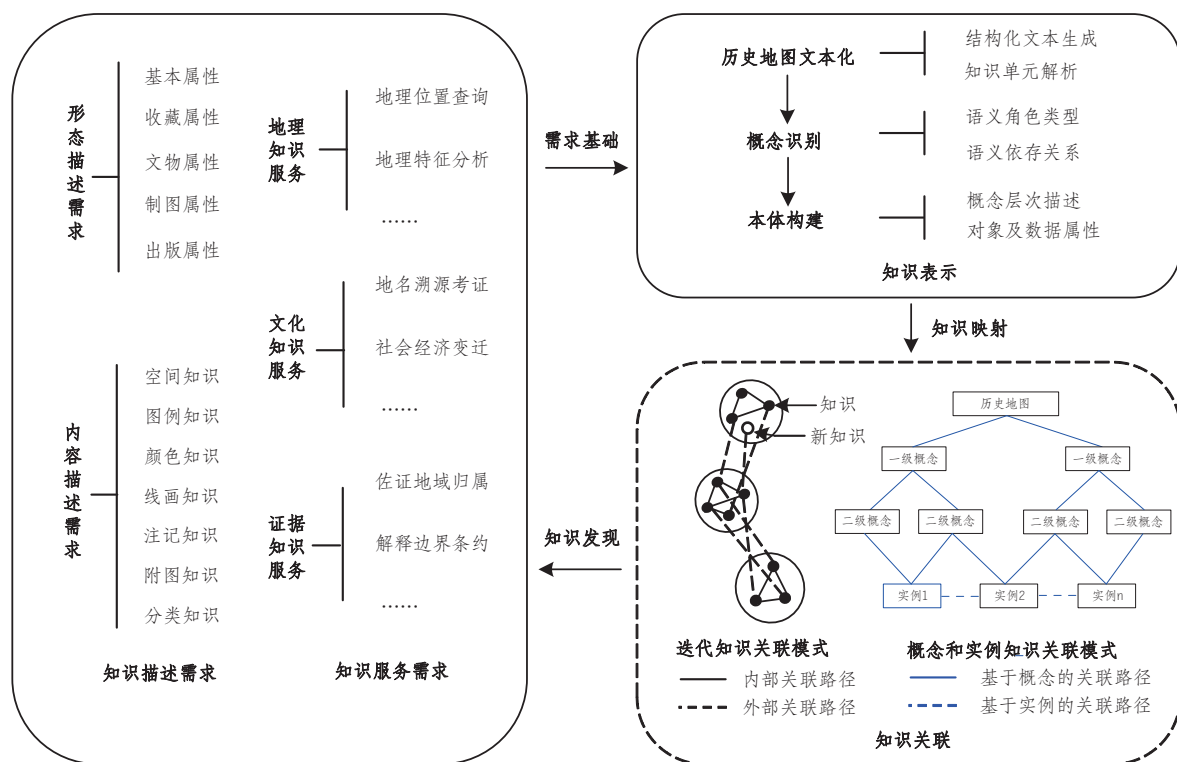


图1 历史地图知识组织框架

2.1 历史地图文本化

针对历史地图的结构化文本生成,本文提出了由文本结构与生成方法组成的一套文本生成规则,如图2所示。该规则的基本原理是:首先,构造基于地图解读文本特征的历史地图结构化文本结构,以保证生成的文本具有完整的信息结构;其次,根据国家颁布的地图领域官方文件与相关地名词典等,构建大规模的历史地图领域语料库,为历史地图结构化文本生成提供科学且规范的地名表达依据;最后,根据历史地图结构化文本结构,采用半自动化方法,即人工与机器协同的方式,对知识描述需求与知识服务需求所涉及的所有信息单元进行融合,最终生成结构化文本,实现历史地图的信息转化。为了确保生成的结构化文本的权威性与准确性,需要通过历史地图领域专家的评估与审核,进行结构化文本的迭代与优化。

其中,“地图形态描述”记录了历史地图获取渠道、收藏信息、制图背景和出版信息等,主要涉及时间、机构、人物、背景以及技术特征等,属于基础语义层次。“地图内容描述”客观记录了历史地图中存在的事物,是对地图空间和元素层级的描述,主要涉及空间、图例、颜色、线画、注记、附图等,这些内容元素相互结合可以初步还原地图的基本结构,属于核心语义层次。“地图价值描述”则代表了历

史地图的延伸价值,即针对历史地图在地理知识服务、文化知识服务、证据知识服务方面所发挥的作用,例如表明、说明、反映、体现、证明等,属于高维语义层次。

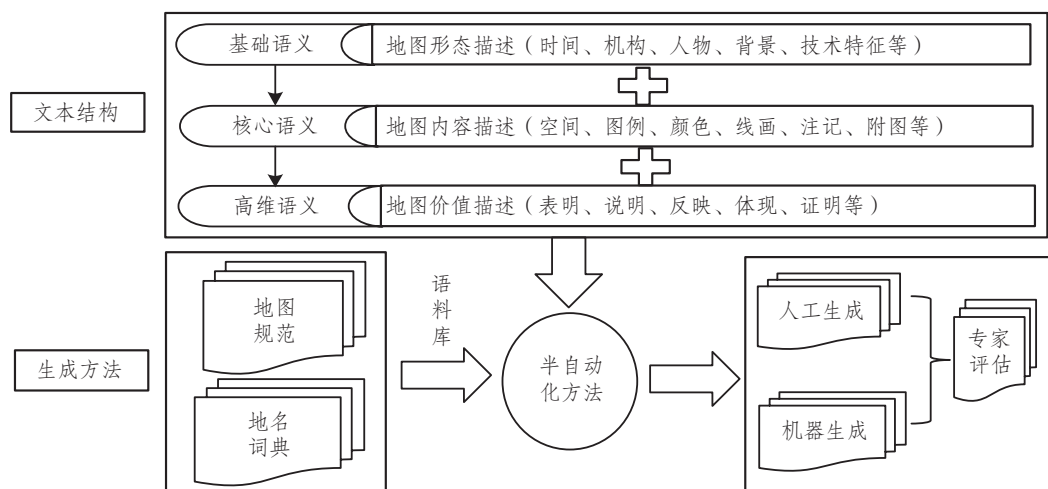


图2 历史地图结构化文本生成规则

本文以国内外图书馆的456幅历史地图作为数据集开展知识组织实践,该数据集共涉及6种语言,7种类型以及多种比例尺,出版时间横跨4个世纪(17—20世纪)。首先,根据知识描述需求与知识服务需求,分别提取456幅历史地图中的形态信息单元、内容信息单元、地理信息单元、文化信息单元、证据信息单元。其次,根据历史地图结构化文本生成规则,采用人工与自然语言处理技术将每幅历史地图的所有信息单元进行融合,生成所有历史地图的结构化文本,经过专家评估与审核之后,对结构化文本进行优化。最后,为了保证句子在结构和语义上的完整性,以句号作为结构化文本的解析依据,对生成的结构化文本进行解析,最终获得1368个有效的句子级别知识单元集。

2.2 概念识别

将句子级别的知识单元解析成核心概念的过程,即概念识别。概念是构成特定领域基本知识框架的语义角色类型与语义依存关系。在语言学领域,语义角色类型是根据句中名词与动词的语义关系而抽象出来的角色类型,包括中心语义角色与外围语义角色^[34]。中心语义角色又包括施事、受事、与事、主事/客事与致事。其中,施事表示行为的主动发出者;受事表示因施事动作行为受到影响的事物;与事为事件的非主动参与者;主事/客事为性质状态发生变化的主体;致事包括时间或变化的引发者。外围语义角色即不是句子必需的成分,主要包括工具、时间、处所等,分别指代动作、行为所凭借的器具、场所与时点。上述各角色类型与领域核心具有一一对应关系。语义依存关系指句子中的“动—名关系”,这些谓词往往阐述角色类型之间的动程、性质与使动关系,起到关联语义角色的作用,是体现语句含义的关键词。因此,通过语义依存分析,可以从知识单元中解析出语义角色类型与语义依存关系。本文使用PyItp依存句法工具包^[35]对1368个知识单元进行语义角色识别和语义依存关系抽取,共获得8种语义角色类型与7类语义依存关系,分别如表3和表4所示。

从表3可知,历史地图的语义角色类型包括时间、参与者、空间、程度、方向、数量、范围、方式。其中,空间、参与者、时间语义角色的频次最高,其次是范围、方向。由此可见,空间、参与者、时间可作为历史地图中的核心概念。从表4中可知,历史地图的语义依存关系包括核心关系、并列关系、先行关

系、后继关系、否定标记、关系标记、依附标记。其中,核心关系包含的语义关系类型最丰富,主要包括标注、绘制、表示、命名等历史地图中的核心关系;并列关系与关系标记主要包括“在……之间、及、和、与、在……内、在……范围、在……附近、在……四周”等用于表达历史地图中空间的关系;先行关系与后继关系主要包括“利用……改绘、以……描述、由……改为、由……组成”等用于表达历史地图中具有的逻辑关系;否定标记与依附标记主要包括“推断……没有、并不是、未说明、不能证明、错误标注、说明了、反映了、阐明了”等用于反映历史地图价值的关系。

表3 历史地图的语义角色类型

角色类型	数 量	标 签	句子级别的知识单元示例
时间	403	TMP	郑光祖于[1839年 TMP]所编撰……
参与者	912	ARG	[陈铎 ARG]编著《广东省明细全图》……
空间	1 241	LOC	店员从[浙江吴兴 LOC]收购……
程度	57	DGR	中国现存[最早 DGR]的一幅全国地图……
方向	79	DIR	画了一条[自东北向西南 DIR]飘荡的风帆……
数量	64	QTY	《郑和航海图》绘制了[109条 QTY]针路航线……
范围	90	EXT	标绘了[我国沿海 EXT]的航线……
方式	48	MNR	用[黑色线条 MNR]标绘航路……

表4 历史地图语义依存关系

关 系	标 签	示 例
核心关系	Root	标注、绘制、表示、命名、公布、出版、记载……
并列关系	eCOO	在……之间、及、和、与……
先行关系	ePREC	利用……改绘、以……描述……
后继关系	eSUCC	由……改为、由……组成、用……标示、采用……标注……
否定标记	mNEG	推断……没有、并不是、未说明、不能证明、错误标注……
关系标记	mRELA	在……内、在……范围、在……附近、在……四周……
依附标记	mDEPD	表明、说明、反映、体现、证明……

3 历史地图本体构建及知识关联

3.1 历史地图本体构建

本体是对某一领域内概念类及类之间关系的形式化表示。通过历史地图资源语义组织的本体模型,使数据的组织形式由粗粒度转变为细粒度,解决历史地图领域知识碎片化、非结构化以及语义化等问题。CIDOC CRM 概念模型声明了 81 个类和 160 个属性,可实现不同来源信息之间的交换和整合,集中应用于文化遗产领域的本体构建、语义揭示、语义融合等方面^[36-39]。

为了进一步将提取得到的概念与关系形成知识网络,本文参考 CIDOC CRM 概念模型,结合上文获得的概念识别结果,设计了历史地图资源语义组织的本体类及其对象属性与数据属性,如图 3 所示。其中,类包括“时间”“空间”“元素”“事件”“参与者”“类型”,分别源于 CIDOC CRM 概念模型中的 E52 Time Span、E53 Place、E70 Thing、E5 Event、E39 Actor、E54 Dimension。根据历史地图的形态知识与内容知识,顶层概念类可细分诸多二级。历史地图本体的对象属性来源于两部分,一部分来自 CIDOC CRM 概念模型,如所属……类型、绘制时间、出版时间、修订时间、有图例、使用……地形表示/

投影、所属……类别等;另一部分来自历史地图的语义依存关系的抽取结果,如有文本、有附图、有注记、有颜色、所属……性质、使用……投影等(见表5)。历史地图本体在参考 CIDOC CRM 概念模型的基础上,结合描述类的基本特征信息,定义其数据属性。例如,“时间”数据属性包括朝代、年份、时期;“参与者”中的人物数据属性包括姓名、职业、职务、贡献。

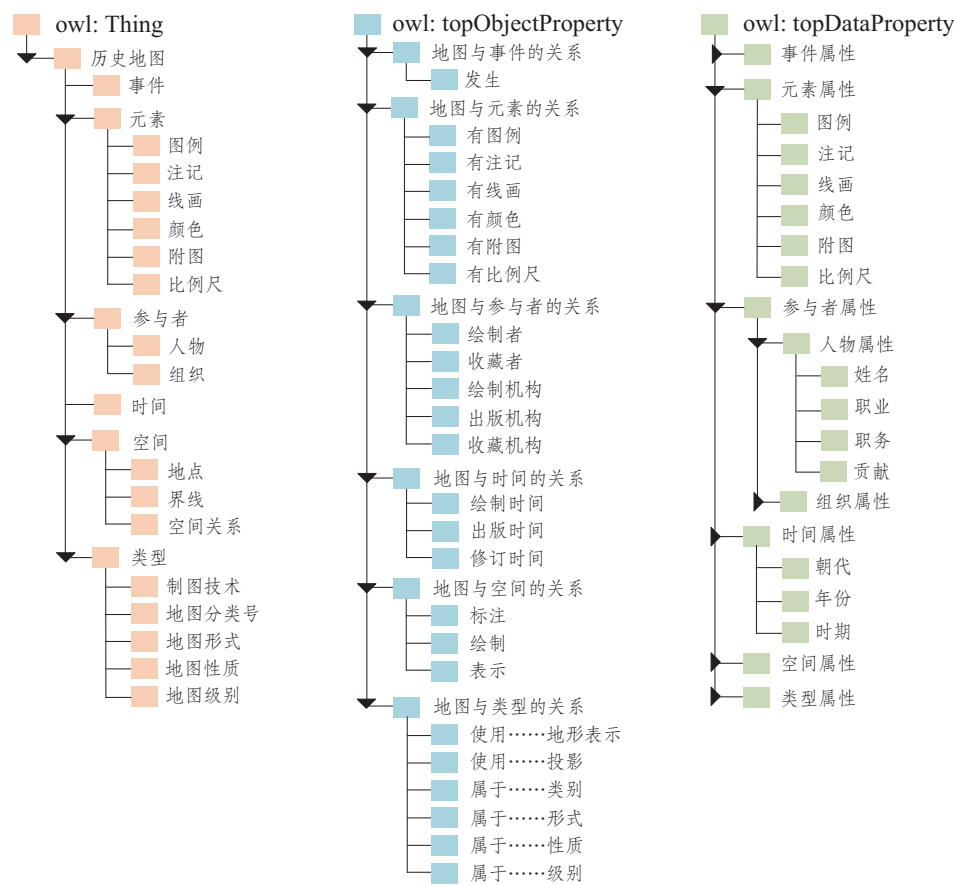


图3 历史地图本体的概念类、对象属性与数据属性

表5 历史地图本体对象属性

	参考对象属性	本文对象属性	定义域	值域
CIDOC CRM 模型	P2 has type	所属……类型	领域地图	类型
	P4 has time-span	绘制/出版/修订时间	领域地图	时间
	P5 consists of	有图例	领域地图	元素
	P32 used general technique	使用……地形表示/投影	领域地图	类型
	P41 classified	所属……类别	领域地图	类型
	P147 curated	收藏	领域地图	参与者
历史地图关系抽取	说明	有文本	领域地图	元素
	附图	有附图	领域地图	元素
	注明;标明	有注记	领域地图	元素
	着色	有颜色	空间	元素
	是;属于	所属……性质	领域地图	类型
	使用	使用……投影	领域地图	类型

3.2 历史地图知识关联

(1) 两种知识关联路径

历史地图的知识关联路径可分为内部知识关联路径和外部知识关联路径。内部知识关联路径指发生在单幅历史地图中的知识关联,是将历史地图内各种符号表征的信息表示成显性知识并实现关联的过程,具体包括单幅历史地图中空间、图例、颜色、线画、注记等元素之间所表达的地理位置、空间关系、行政划分等内容。这些关联信息独立于其他历史地图而存在。外部知识关联路径指发生在历史地图之间的知识关联,是将每幅历史地图看作一个信息集合,并将各个信息集合再次关联以获得新知识的过程。历史地图内外部知识关联路径相结合可以打破“信息孤岛”,是历史地图知识关联实现从量变到质变的基础。

(2) 两种知识关联模式

为了更好地形成历史地图知识网络,本文在传统实例关联的基础上引入基于概念的知识关联方法。基于实例的关联是当两个实例相同或相关时则建立两者之间的关联,但当实例之间的表达方式与语义相距较远时则较难建立关联关系。基于概念的关联首先将具体实例进行概念识别并与本体概念建立关联,随后基于本体关联各概念与实例,建立“实例—概念—实例”的关联关系,使关联层级更丰富,能够满足更加复杂的知识发现需求。若实例1与实例2之间由于表达差异无法直接产生基于实例的关联时,二者可通过概念之间的关系建立起关联路径。基于实例的关联和基于概念的关联的结合可以实现多角度、多层次的知识关联。“多角度”即实例之间存在的多种关联,是横向的关联;“多层次”即概念之间的上下位关系,是纵向的关联。此外,在知识关联过程中会产生新的知识,新的语义关系将作为新的知识增加到历史地图的知识库中以实现语义增强,形成迭代更新机制,实现从历史地图的知识发现到知识涌现。

4 讨论与结论

本文围绕历史地图的知识描述需求和知识服务需求,建立了以知识表示和知识关联为路径的历史地图知识组织框架。知识表示是将单幅历史地图的抽象信息转换为知识单元的过程,具体包括历史地图文本化、概念识别以及本体构建环节。知识关联强调单幅或多幅历史地图共同呈现的综合知识,具体包括内部关联与外部关联两种关联路径以及基于概念与基于实例的两种关联模式。未来将进一步融合与历史地图相关的传统资料(如历史文献、图片、音频以及视频等)与新型资源(如网络资源、口述历史、社交媒体、关联数据、知识图谱、VR等)^[40],通过多模态知识关联实现历史地图内容的全面揭示和分析,以提供更加立体化、个性化、智能化的历史地图知识服务。

在新文科建设与数字人文的背景下,图情档学科迎来了良好的发展契机。图情档学科本身是实践性极强的学科,我国图情档学科应当面向领域前沿最迫切、最重要、最具价值的研究课题,大力投入学术与人才资源,在保障学科学术价值的同时最大程度实现其实践意义^[41]。历史地图的知识组织是一个典型的跨学科研究,涉及历史学、地理学、文学、文献学、目录学、图书馆学、情报学、档案学、出版学、计算机科学,历史地图的结构化、数据化、语义化研究是对图情档学科传统优势领域的创新与延伸。因此,对于历史地图的知识组织,需要图情档学科集中学术资源与力量,以“信息资源管理”和“信息组织”的理论框架和方法为基础,实现跨学科研究的融合创新,提高我国图情档学科的影响力。

致谢:本文系国家社会科学基金重大项目“南海疆文献资料整理中的知识发现与维权证据链建构研究”(项目编号:19ZDA347)和南京大学研究阐释党的二十大精神专项课题“建构中国自主的南海维权知识体系”的研究成果。

参考文献

- [1] 华林甫. 110年来中国历史地图集的编绘成就与未来展望[J]. 中国历史地理论丛, 2021, 36(3): 110-125. (HUA L F. The achievements and future prospects of research and compilation of China's historical atlas in the past 110 years[J]. Journal of Chinese Historical Geography, 2021, 36(3): 110-125.)
- [2] 徐绪堪, 蒋勋, 苏新宁. 面向知识服务的知识组织框架体系构建[J]. 情报学报, 2013, 32(12): 1278-1287. (XU X K, JIANG X, SU X N. Knowledge service-oriented knowledge organization framework and system[J]. Journal of the China Society for Scientific and Technical Information, 2013, 32(12): 1278-1287.)
- [3] 谢秋学, 雍志娟, 贺芳, 等. 构建专业领域知识服务体系的通用框架[J]. 出版发行研究, 2018(8): 35-37. (XIE Q X, YONG Z J, HE F, et al. A general framework for building a professional domain knowledge service system[J]. Publishing Research, 2018(8): 35-37.)
- [4] 潘刚, 张运良, 钟庆虹. 工程科技领域知识服务的思考与实践[J]. 情报工程, 2018, 4(5): 4-12. (PAN G, ZHANG Y L, ZHONG Q H. Thinking and practice of knowledge service in engineering field[J]. Technology Intelligence Engineering, 2018, 4(5): 4-12.)
- [5] 秦春秀, 郑梦悦, 马续补, 等. 基于智能主题图的科技文献细粒度知识组织模型[J]. 情报杂志, 2021, 40(11): 169-175. (QIN C X, ZHENG M Y, MA X B, et al. A fine-grained knowledge organization model for scientific literature based on intelligent topic graphs[J]. Journal of Intelligence, 2021, 40(11): 169-175.)
- [6] 庄文杰, 谈国新, 侯西龙, 等. 非物质文化遗产视频知识元组织模型研究[J]. 情报科学, 2018, 36(12): 25-32. (ZHUANG W J, TAN G X, HOU X L, et al. The organization model of intangible cultural heritage video knowledge element[J]. Information Science, 2018, 36(12): 25-32.)
- [7] 程结晶, 王心雨. 敦煌遗书图像知识关联与语义描述[J]. 图书情报工作, 2021, 65(7): 123-131. (CHENG J J, WANG X Y. Knowledge association and semantic description of Dunhuang documents images[J]. Library and Information Service, 2021, 65(7): 123-131.)
- [8] 周知, 蒋琳. 数字人文图像资源知识组织模型构建研究[J]. 图书馆学研究, 2021(8): 66-72. (ZHOU Z, JIANG L. Research on the construction of knowledge organization model of digital humanistic image resources[J]. Research on Library Science, 2021(8): 66-72.)
- [9] USERY E L, VARANKA D. Design and development of linked data from the national map[J]. Semantic Web, 2012, 3(4): 371-384.
- [10] STADLER C, LEHMANN J, HÖFFNER K, et al. LinkedGeoData: a core for a web of spatial open data[J]. Semantic Web, 2012, 3(4): 333-354.
- [11] LI H L, LIU J, ZHOU X R. Intelligent map reader: a framework for topographic map understanding with deep learning and gazetteer[J]. IEEE Access, 2018, 6: 25363-25376.
- [12] JANOWICZ K, KEBLER C. The role of ontology in improving gazetteer interaction[J]. International Journal of Geographical Information Science, 2008, 22(10): 1129-1157.
- [13] ZHOU X R. GeoAI-enhanced techniques to support geographical knowledge discovery from big geospatial data[D]. Tempe: Arizona State University, 2019.
- [14] BATTLE R, KOLAS D. Enabling the geospatial semantic web with parliament and GeoSPARQL[J]. Semantic Web, 2012, 3(4): 355-370.

- [15] 周熙然,李德仁,薛勇,等. 地图图像智能识别与理解:特征、方法与展望[J]. 武汉大学学报(信息科学版), 2022,47(5):641-650. (ZHOU X R, LI D R, XUE Y, et al. Intelligent map image recognition and understanding: representative features, methodology and prospects[J]. Geomatics and Information Science of Wuhan University, 2022,47(5):641-650.)
- [16] 肖珑,苏品红,姚伯岳. 国家图书馆舆图元数据规范与著录规则[M]. 北京:国家图书馆出版社,2014:34-43. (XIAO L, SU P H, YAO B Y. Metadata specifications and description rules of maps in the National Library of China [M]. Beijing: National Library of China Publishing House, 2014:34-43.)
- [17] KUHN W. Approaching the issue of information loss in geographic data transfers[J]. Geographical Systems, 1997,4(3):261-276.
- [18] 田江鹏,游雄,贾奋励,等. 地图符号的认知语义分析与动态生成[J]. 测绘学报,2017,46(7):928-938. (TIAN J P, YOU X, JIA F L, et al. Cognitive semantic analysis and dynamic generation of cartographic symbols[J]. Acta Geodaetica et Cartographica Sinica, 2017,46(7):928-938.)
- [19] 杜清运. 空间信息的语言学特征及其自动理解机制研究[D]. 武汉:武汉大学,2001. (DU Q Y. Study on linguistic characteristics and automatic understanding mechanism of spatial information[D]. Wuhan: Wuhan University, 2001.)
- [20] 陈亮明. 作图之准备[J]. 水陆地图审查委员会会刊,1935(1):9-15. (CHEN J M. Preparation for drawing[J]. Journal of Land and Water Map Review Committee, 1935(1):9-15.)
- [21] 王晓光,梁梦丽,侯西龙,等. 文化遗产智能计算的肇始与趋势——欧洲时光机案例分析[J]. 中国图书馆学报,2022,48(1):62-76. (WANG X G, LIANG M L, HOU X L, et al. Intelligent computing of cultural heritage: a case study of European Time Machine project[J]. Journal of Library Science in China, 2022,48(1):62-76.)
- [22] 慎利,徐柱,李志林,等. 从地理信息服务到地理知识服务:基本问题与发展路径[J]. 测绘学报,2021,50(9):1194-1202. (SHEN L, XU Z, LI Z L, et al. From geographic information service to geographic knowledge service: research issues and development roadmap[J]. Acta Geodaetica et Cartographica Sinica, 2021,50(9):1194-1202.)
- [23] 侯涛. 地理分析模型知识图谱构建与分析方法研究[D]. 南京:南京师范大学,2019. (HOU T. Research on knowledge map construction and analysis method of geographic analysis model[D]. Nanjing: Nanjing Normal University, 2019.)
- [24] EGENHOFER M J. Toward the semantic geospatial web[C]//Proceedings of the 10th ACM International Symposium on Advances in Geographic Information Systems. New York: Association for Computing Machinery, 2002:1-4.
- [25] 王利兵. 地图与话语:海洋边界建构的国家实践[J]. 云南师范大学学报(哲学社会科学版),2019,51(6):9-17. (WANG L B. Map and discourse: national practice of marine boundary construction[J]. Journal of Yunnan Normal University (Humanities and Social Sciences Edition), 2019,51(6):9-17.)
- [26] 1717年清朝官方绘制的南海地图[EB/OL]. [2022-05-14]. http://www.thesouthchinasea.org.cn/2016-07/18/c_53206.htm. (An official map of the South China Sea drawn by the Qing Dynasty in 1717[EB/OL]. [2022-05-14]. http://www.thesouthchinasea.org.cn/2016-07/18/c_53206.htm.)
- [27] 司徒尚纪. 中国南海海洋文化[M]. 广州:中山大学出版社,2009:102-111. (SITU S J. South China Sea marine culture[M]. Guangzhou: Sun Yat-sen University Press, 2009:102-111.)
- [28] 徐丽坤,徐汉卿,张海静,等. 对湖北省地图整理应用的研究和思考[J]. 地理空间信息,2020,18(2):56-58,7. (XU L K, XU H Q, ZHANG H J, et al. Research and thinking on the application of map sorting in Hubei Province [J]. Geospatial Information, 2020,18(2):56-58,7.)
- [29] 苏品红. 地图资源的开发与利用初探[J]. 中国图书馆学报,1998,24(3):82-85. (SU P H. On the development and use of map resources[J]. Journal of Library Science in China, 1998,24(3):82-85.)

- [30] 郑志华,吴静楠.《中国南海各岛屿图》的国际法意义探微——兼对比尔·海顿的回应[M]//中国海洋法年刊(2020).北京:知识产权出版社有限责任公司,2020:111-119. (ZHENG Z H, WU J N. Exploring the international legal significance of the *Maps of Islands in the South China Sea* and its response to Bill HAYDN[M]//Chinese Yearbook of the Law of the Sea 2020. Beijing: Intellectual Property Publishing House Co., Ltd, 2020: 111-119.)
- [31] 余元元. 地图在领土及边界争端解决中的作用:国际法院判例研究[D]. 广州:中山大学,2011. (SHE Y Y. The role of maps in the settlement of territorial and boundary disputes in the jurisprudence of the ICJ[D]. Guangzhou: Sun Yat-sen University, 2011.)
- [32] GKADOLOU E, TOMAI E, STEFANAKIS E, et al. Ontological standardization for historical map collections: studying the Greek borderlines of 1881[C]//ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Melbourne, Australia, 2012, 1-2: 203-208.
- [33] 洪亮,马费成. 面向大数据管理决策的知识关联分析与知识大图构建[J]. 管理世界, 2022, 38(1): 207-219. (HONG L, MA F C. Knowledge association analysis and big knowledge graph construction for big data management and decision-making[J]. Management World, 2022, 38(1): 207-219.)
- [34] 文旭. 语言学导论[M]. 北京:北京师范大学出版社, 2013: 3. (WEN X. An introduction to linguistics[J]. Beijing: Beijing Normal University Press, 2013: 3.)
- [35] CHE W X, LI Z H, LIU T. LTP: a Chinese language technology platform[C]//Proceedings of the 23rd International Conference on Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2010: 13-16.
- [36] 戴畋. 基于 CIDOC CRM 的博物馆文物数字化影像元数据规范研究——以中国国家博物馆文物影像元数据体系设计为例[J]. 中国博物馆, 2020(3): 131-136. (DAI T. Research on metadata specification of digital image of museum cultural relics based on CIDOC CRM: taking the design of cultural relics image metadata system of National Museum of China as an example[J]. Chinese Museum, 2020(3): 131-136.)
- [37] DAVIDE V, DORA M, PIMENTA R I. A tool to explore the population of a CIDOC-CRM ontology[J]. Procedia Computer Science, 2021(192): 158-167.
- [38] 侯西龙,谈国新,庄文杰,等. 基于关联数据的非物质文化遗产知识管理研究[J]. 中国图书馆学报, 2019, 45(2): 88-108. (HOU X L, TAN G X, ZHUANG W J, et al. Research on knowledge management of intangible cultural heritage based on linked data[J]. Journal of Library Science in China, 2019, 45(2): 88-108.)
- [39] HYVÖNEN E. Digital humanities on the semantic web: sampo model and portal series[J]. Semantic Web, 2023, 14(4): 729-744.
- [40] 魏大威,李志尧,刘晶晶,等. 基于区块链技术的智慧图书馆数字资源管理研究[J]. 中国图书馆学报, 2022, 48(2): 4-12. (WEI D W, LI Z Y, LIU J J, et al. Digital resource management of smart library based on blockchain technology[J]. Journal of Library Science in China, 2022, 48(2): 4-12.)
- [41] 张洋,朱嘉麒,王媛媛,等. 图情学科研究的范式思考——以科学评价为例[J]. 图书情报工作, 2021, 65(12): 20-26. (ZHANG Y, ZHU J Q, WANG Y Y, et al. Reflections on the research paradigm in library & information science: a case study of scientific evaluation[J]. Library and Information Service, 2021, 65(12): 20-26.)

齐小英 南京大学信息管理学院博士研究生。江苏 南京 210023。

杨海平 南京大学信息管理学院教授,南京大学出版研究院副院长。江苏 南京 210023。

(收稿日期:2022-07-24;修回日期:2024-03-01)