

文化大数据背景下线性文化遗产资源 知识表示及语义增强框架构建

孙绍丹 张莉曼 朱 祥

[摘 要] [目的/意义]构建线性文化遗产资源知识表示及语义增强框架,为文化机构开展文化遗产智慧化数据建设及推动优秀传统文化创新性发展提供路径参考。[方法/过程]在提炼线性文化遗产“资源族群性、时间流动性和空间链状性”特征基础上,剖析线性文化遗产知识体系的层级结构,建构线性文化遗产资源知识表示模型,并依据表征理论和情境认知理论,从结构化语义增强和情境化语义增强视角对线性文化遗产数据进行语义增强,创新性地提出“时空”双重维度语义增强路径,最后以京杭大运河线性文化遗产为例,进行大运河知识表示及语义增强案例研究。[结果/结论]本文构建的知识表示模型及语义增强框架有利于推进线性文化遗产数据的多维语义组织和深度挖掘研究,提升线性文化遗产内容的语义互联性和外延性。

[关键词]文化大数据;线性文化遗产;知识表示;语义增强;京杭大运河

[作者简介]孙绍丹(1993-),女,南京理工大学网络空间安全学院讲师,博士,研究方向:知识组织,语义增强;张莉曼(通讯作者)(1994-),女,南京理工大学网络空间安全学院讲师,博士,硕士生导师,研究方向:知识表示(江苏 南京 210094);朱祥(1992-),男,南京理工大学网络空间安全学院讲师,博士,研究方向:知识融合。

[原文出处]《现代情报》(长春),2023.11.96~111

[基金项目]江苏省高校哲学社会科学研究一般项目“面向智慧服务的公共数字文化数据语义增强研究”(项目编号:2023SJYB0032);教育部人文社会科学青年基金项目“面向决策支持的多源政民互动数据语义融合及故事化呈现研究”(项目编号:22YJC870020);江苏省高校哲学社会科学研究一般项目“基于多源异构知识网络和深度学习的药物知识发现研究”(项目编号:2023SJYB0016)。

2022年,“国家文化数字化战略”意见明确提到“十四五”时期末,基本建成物理分布、逻辑关联、快速链接、高效搜索、全面共享、重点集成的国家文化大数据体系,使中华文化全景呈现。由此可见,文化大数据浪潮已然迭起。线性文化遗产是指拥有特殊文化资源集合的线性或带状区域内的物质和非物质文化遗产族群,如运河、廊道、峡谷、古道等,其理念是通过沿对沿线文化遗产进行有效串联,构成链状的文化遗存状态,形成长期赓续的遗产生态系统,实现遗产的整体化保护与活化^[1]。作为重要的文化资源,线性文化遗产蕴含着丰富的人文意义和文化内涵,亟待依托大数据和人工智能技术深度组织和挖掘遗产内容要素,钩沉线性文化遗产中的精神之源和思想精髓,让遗产中的厚重文化基因与当代形成价值和鸣,从而推

动中华优秀传统文化的创造性转化和创新性发展。

本文通过提炼线性文化遗产资源的时空属性特征,构建线性文化遗产知识表示模型。基于认知心理学中的表征理论和情境认知理论,提出从结构化语义增强和情境化语义增强两个维度对线性文化遗产进行语义增强,并以京杭大运河为例,对线性文化遗产知识表示模型进行实例研究,在此基础上进行结构化和情境化语义增强,从而提升线性文化遗产内容的语义互联性和外延性,推动线性文化遗产资源共享传播和活化利用。

1 研究现状

1.1 线性文化遗产研究现状

通过文献调研,国内外有关线性文化遗产的研究呈现多学科分布特点,如地理学、环境科学、历史

学、艺术学、传播学、图情档等学科都对线性文化遗产进行不同视角的探索。研究内容聚焦于线性文化遗产价值评估、旅游开发、活化利用模式构建、遗产廊道构建、保护模式探索等方面,如探讨虚拟现实及数字叙事技术在线性文化遗产旅游开发中的应用^[2]、提出线性文化遗产的文化价值观塑造^[3]、探索线性文化遗产沿线重要节点旅游发展模式^[4]、构建分层次、有重点的线性文化遗产保护普适性空间格局^[5]。在项目实践方面,北京大学俞孔坚教授牵头构建了“国家线性文化遗产网”,提出涵盖交通线路、军事工程、自然河流与水利工程及历史主题事件四大类型的19个线性文化遗产^[6],如大运河、丝绸之路、长城、茶马古道等。在线性文化遗产资源组织研究方面,国内外成果相对较少,如徐芳教授以潇贺古道和大运河线性文化遗产为例,在元数据描述和关联数据模型构建方面发表了系列成果^[7],但局限于资源的浅层组织层面,缺乏对线性文化遗产特征的深层剖析,尚未对资源内容进行细粒度解构。

1.2 知识表示研究现状

知识表示是指在计算机系统中对领域知识进行抽象、表达和存储的一种方式,一般采用逻辑表示、框架表示、语义网络及本体等形式化工具来表示知识,以便计算机能够理解、推理和处理。在文化遗产领域,知识表示相关研究主要以理论探索和模型构建为聚焦点,研究对象聚焦于文物、文化遗产图像、印章、历史报纸文献^[8]、方志物产及非遗资源等。在理论探索方面,主要以知识组织为研究核心,探讨文化遗产领域语义组织研究现状^[9],剖析文化遗产领域核心共享概念及领域知识关联关系,构建文化遗产领域知识纲要^[10]。在知识表示模型构建研究方面,主要以本体和知识图谱构建为技术线路^[11],构建文化遗产知识表示模型。如Wang X等^[12]采用知识图谱方法对文化遗产进行知识表示,借助机器学习技术和语义Web技术,将元数据中的实体、属性和关系与外部知识资源进行链接,从而建立文化遗产知识图谱。宋雪雁等^[13]采用知识图谱技术对清代祭祀礼器进行知识表示、关联及发现研究。张卫等^[14]构建了非遗“古琴艺术”细粒度知识语义表示及关联模型。同时,针对叙事型文化遗产,有学者提出基于角色关

联的知识表示方式,以解决现有知识表示方法存在的角色缺乏情境依赖与表示角度单一问题。

1.3 语义增强研究现状

语义增强是伴随语义网技术和文本处理技术而逐渐发展起来的概念,指采用各种技术和方法向内容添加语义元数据,旨在增强数据的语义理解力^[15],与语义描述、语义标注、语义分析、语义关联、语义融合和语义发现等知识组织活动息息相关。欧盟数字图档博联盟也构建了语义增强框架,将其划分为分析、链接和扩充3个阶段。其中分析阶段着重于元数据框架的制定,定义原始字段与语境资源之间的匹配与链接规则;链接阶段是采用实体链接等方式自动将元数据字段的值与语境资源的值进行匹配,从而将语境链接添加到数据集当中;扩充阶段是指通过外部数据源或知识库等方式实现原始对象资源语义的增强。Damjanovic-Behrendt V等^[16]将语义增强过程划分为语义搜索、语义中介、语义标注、语义分析和知识发现阶段。语义搜索阶段旨在用语义技术增强传统关键字搜索,提高查全率和查准率,如进行多面检索、语义自动补全等;语义中介阶段包括数据合并和映射,合并主要指重叠部分的本体之间的合并行为。映射是构建映射语句,定义本体间转换规则;语义标注阶段旨在明晰文档中概念及概念之间的关系,使机器可理解和可互操作;语义分析和知识发现阶段是通过分析实体语义关系来进行数据搜索和知识发现的过程。近年来,在文化遗产领域,语义增强技术逐渐受到关注。如Abgaz Y等^[17]基于人工智能技术对文化遗产图像资源进行语义增强,采用深度学习模型进行图像识别和标注,并用本体论方法对图像元数据进行描述和分类,提高了图像的可访问性和可发现性。Tan X等^[18]基于图像学理论,提出文化遗产数字图像语义丰富的系统方法和技术途径,如语义注释、基于实体的语义增强、建构数据关联关系、以事件为中心的语义增强、定义属性之间的层级关系、命名实体识别等,以此增强文化遗产图像的语义化特征。

综上,文化遗产领域的知识表示及语义增强研究成果相对丰富,但针对线性文化遗产资源进行知识组织、知识表示及语义增强的文献相对不足,且缺

乏对线性文化遗产资源特征进行分析提炼,区别于点状或面状的遗产资源,线性文化遗产往往呈现“遗产族群性、空间链状性、时间流动性”等特征,如何结合上述特征,进行线性文化遗产资源的知识表示及语义增强研究,成为学界关注焦点。鉴于此,本文通过构建线性文化遗产资源知识表示及语义增强框架,以大运河为例,旨在为线性文化遗产资源的语义组织提供可操作性的方法指导,推进线性文化遗产资源的开发和利用。

2 线性文化遗产资源特征分析

线性文化遗产一词是由原故宫博物院院长单霁翔先生提出,强调遗产资源的链状特征和集群效应。截至2023年3月,联合国教科文组织世界遗产中心发布的《世界遗产名录》中,共有1157处世界遗产^[19],主要以自然遗产、文化遗产和双重遗产进行分类。其中,中国世界遗产总数为56处。虽然国内外学者对世界遗产类型进行研究,但尚未剥离出线性文化遗产资源类型。张书颖等^[20]结合世界遗产委员会的遗产项目评语,根据《世界遗产名录》总结出33处线性文化遗产,中国有代表的当数长城、大运河以及丝绸之路—长安—天山走廊路线。这些资源文化和历史价值丰富,亟待现代技术赋能推动资源的活化利用。本文参考文献[20]对线性文化遗产资源组合模式分析,如图1所示,结合线性文化遗产的概念及内涵,以京杭大运河为例,阐释线性文化遗产资源的主要特征。

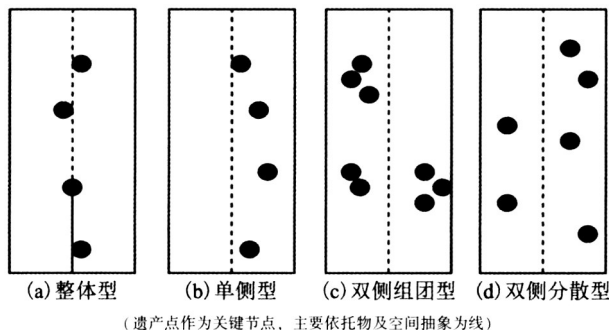


图1 线性文化遗产资源组合模式

2.1 资源族群性

在线性文化遗产资源形成和发展过程中,人类活动的群体性质一定程度上塑造了线性文化遗产的族群性特征,沿线地域的人文地理特征和历史文化

积淀成独特的文化遗产资源,分布在一定的地域范围内,因此族群性是线性文化遗产最显著的特征之一。由图1可以看出,资源点聚集性分布,沿中心轴线呈单侧或双侧分布。例如京杭大运河是荟萃沿线众多物质和非物质文化遗产的线性廊道,包括物质文化遗产(如扬州瘦西湖、苏州盘门等)和非物质文化遗产(如锡剧、昆曲等)。相对于单点的遗产资源,线性文化遗产集聚性特征更加突出,从单一范围的“点”转向“线”“面”,经过自然、历史、社会和文化等多方面因素的影响,形成独具特色的资源集群。

2.2 空间链状性

由图1可以看出,线性文化遗产在空间分布上呈现“链状”性特征,具有明显的空间延伸性和连续性,既反映了地域文化的独特性和复杂性,也反映了社会、政治、经济和文化等因素的交织和影响。在线性文化遗产空间结构研究方面,李飞^[21]提出线性文化遗产空间要素包括节点、廊道和辐射域面。节点即资源点,节点之间相互作用影响形成了廊道,节点和廊道在自然和社会环境共同作用下形成了辐射域面。整体来看,线性文化遗产空间结构也随着社会变革和环境影响进行不断的演化,如京杭大运河在不同历史时期河道变迁历程复杂多样,受文化、政策、军事、经济等多种因素影响,但总体上呈现出链状分布特征,在地理空间上表现出明显的延伸性。因此,在对线性文化遗产资源进行知识表示及语义增强时,需要充分考虑其空间链状性特征,将空间因素考虑到语义增强框架构建过程中。

2.3 时间流动性

时间流动性指的是线性文化遗产在历史长河中的流动和演变特征。具体来说,这些遗产的产生、发展和消亡都是在特定的历史时期和历史背景下发生的,需要历经时间的沉淀和积累,记录了人类社会在不同历史时期的文化、经济、政治和社会变迁,是历史文化发展的产物,呈现出动态变化特征。例如京杭大运河历史悠久,溯源于春秋,完成于隋朝,繁荣于唐宋,取直于元,疏通于明清,距今已2500多年的历史。大运河的开凿、发展及繁荣的历史进程凝结着中国人民的智慧。不管是因河道、船闸、行船等而兴起的文物资源,还是沿线的戏曲、文学、艺术、美

食、民谣等非遗资源都在流动的变迁之中,孕育着中华民族顽强不息的文化基因。放眼于历史长河,大运河线性遗产尽显历史沧桑变迁的年轮,具有典型的时间流动性特征。因此,在对线性文化遗产资源进行知识表示及语义增强时,也同样需要考虑资源的时间序列关系,构建时空整合的语义模型,增强数据的语义特征。

3 线性文化遗产资源知识表示及语义增强研究框架

本文在线性文化遗产资源特征分析基础上,构建线性文化遗产知识表示及语义增强框架,如图2所示。首先,从知识概念本源视角出发,解析知识的分类维度,从陈述性知识、程序性知识和情境性知识3个维度进行定义,结合线性文化遗产资源特征,分析其知识体系的层级结构;其次,采用元数据、本体建模技术,参考成熟的本体模型,从概念对象层、内容表征

层和语义情境层构建线性文化遗产知识表示模型;最后,以认知科学中的表征理论和情境认知理论为基础,从结构化语义增强和情境化语义增强维度探究线性文化遗产语义增强方法。在上述框架基础上,以京杭大运河线性文化遗产为例,进行大运河知识体系构建、知识表示模型构建及语义增强方法研究。

3.1 线性文化遗产知识体系层级结构

知识是人们在改造世界的实践中所获得的认识和经验的总和,是对客观世界的概括和总结,也是对自然界、人类社会和思维活动规律的认识。知识体系是描述特定专业知识综合的概念性术语,知识需要人们经过实践活动获取经验,对经验进行消化、积累、思考、归纳、理解、抽象而上升为知识体系。本文参考经济合作与发展组织(OECD)对知识分类,并结合用户情境认知理论中对于知识的获取要依赖一定的社会情境,归纳出陈述性知识、程序性知识和情境性知识3个维度,以此为基础,在不同的知识维度内定义知识体系的层级结构,如图3所示。

陈述性知识包括概念层和对象层,程序性知识包括状态层和特征层,情境性知识包括关系层、时空层和叙事层。概念层主要用以表达线性文化遗产知识体系中的概念描述,从宏观维度解析,将线性文化遗产划分为物质文化遗产和非物质文化遗产,前者又可以进一步划分为可移动文物和不可移动文物,概念结构层层递进;对象层主要用以表示线性文化遗产的构成对象;状态层是指对象在特定的事件演化过程中所发生的变化、动作和记录;特征层定义了不同对象所具有的外部特性;关系层用以表达线性文化遗产概念、对象、状态等之间的联系;时空层主要以建构时空框架为主要目标,参考俞肇元等^[22]学

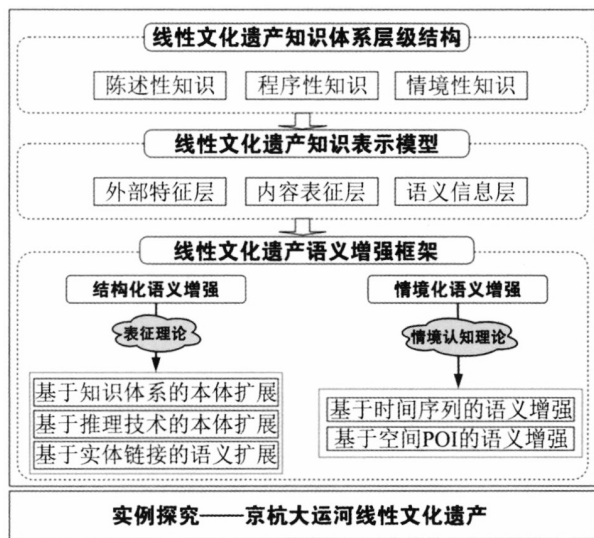


图2 研究框架

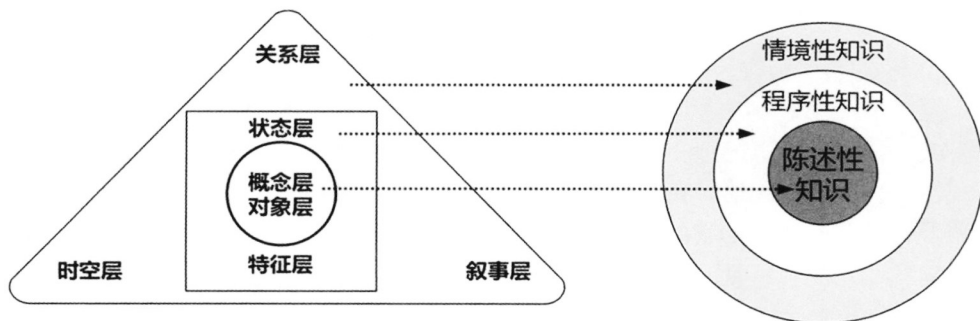


图3 知识体系的层级结构

者提出的地理信息要素“七维度”,即地理语义、空间位置、几何形态、属性特征、要素关系、演化过程和作用机制,可以构建地理信息时空框架图;叙事层是以数据故事化理论为基础,对叙事过程进行建模分析,遵循数据故事工程化思维展开研究。

3.2 线性文化遗产知识表示模型

知识表示是一种用来描述和组织知识的形式化方法,通过定义实体及其关系来表示知识。线性文化遗产知识表示模型是一种用于描述和组织线性文化遗产数据的形式化方法,旨在将复杂的线性文化遗产知识以计算机可理解和处理的形式进行表示和存储。本文在3.1知识体系层级结构基础上,将概念层和对象层概括为概念对象层,将状态层和特征层概括为内容表征层,将关系层、时空层及叙事层概括为语义情境层,从以上3个维度构建线性文化遗产资源知识表示模型。

概念对象层聚焦于线性文化遗产知识体系中的概念描述和对象表达,将概念和对象抽象为本体建模中的类和实例,定义概念类和对象实例之间的层次关系。内容表征层聚焦于线性文化遗产的内容特征及状态变化,将特征和状态抽象为本体中的数值属性和对象属性,以表示线性文化遗产对象的特征和状态,如文物的年代、材质、尺寸等。语义情境层聚焦于线性文化遗产内容的语义关系及语境信息的传递及表达,如所属文化机构及其在时间和空间上的关联,实现对知识的关联性分析和语义丰富。

3.3 线性文化遗产语义增强框架

语义增强本质上是一种语义组织活动,能进一步揭示数据语义内涵,增强数据内容表达的外延能力,提高数据利用价值。本文在线性文化遗产知识表示模型基础上,结合线性文化遗产“资源族群性、空间链状性、时间流动性”等特征,从认知心理学的表征理论和情境认知理论解析语义增强的两个不同维度,即结构化语义增强和情境化语义增强。

表征理论是指人类思维中的信息被组织成一系列表征,每个表征均包含事物的知识和特征,任何表征都依赖一定的结构,即表征方式或结构方式。结构化语义增强从表征理论出发,在原有知识体系基础上,对结构化数据进行语义扩展,如基于本体扩展

的语义增强和基于实体链接的语义增强等,目的是丰富数据的语义表达,更好地支持数据复用、共享和智能化处理。线性文化遗产结构化语义增强是从资源特征信息入手,在原有数据基础上增加有关此数据的外部信息源,构建资源的“聚集性”和“网络性”集群,契合线性文化遗产的族群性特征,实现数据的互联互通。

情境认知理论是在行为主义、认知学习理论及建构主义基础上发展起来的,强调知识获取依赖一定的情境,不能孤立存在。情境化语义增强考虑情境认知理论相关要素,结合线性文化遗产的时空特性,从时间和空间两个维度建构知识的情境场域,拓展线性文化遗产语义表达内涵,让语义描述更加丰富、精准,以提高信息准确性和实用性,满足用户个性化和沉浸化需求。线性文化遗产情境化语义增强是从线性文化遗产语义信息入手,在原有数据基础上增加多维信息特征和内容,结合线性文化遗产资源的空间链状性和时间流动性特征,采用时间序列方法及空间POI(Point of Interest)技术来构建线性文化遗产的“情境化”集群,实现数据集成共享。

4 实例探究——京杭大运河线性文化遗产知识表示及语义增强

4.1 京杭大运河线性文化遗产概览

京杭大运河是世界上最长的人工运河,始建于春秋时期,见证了我国2500余年的历史,北起通州,南至杭州,流经北京、天津、河北、山东、江苏、浙江等省市,全长1794km,是中国历史上南粮北运、军资调配和水利灌溉等用途的生命航道,是中华民族的宝贵财富。大运河沿线文化遗产资源丰富,沿河兴起的码头、城镇、船闸、桥梁、堤坝、石刻、墓葬、建筑园林、戏曲歌舞等形成了大量的线性文化遗产,其中拥有世界文化遗产的河道遗产、水工遗存及相关遗产共计85处,非物质文化遗产450余项^[23]。2014年6月,京杭大运河被正式列入《世界文化遗产名录》。2021年8月,国家文化公园建设工作领导小组印发《大运河国家文化公园建设保护规划》,以整合大运河沿线8个省市文物和文化资源为主要目标。2022年6月,江苏省文投集团打造了大运河国家文化公园数字云平台,由“知运河、探运河、游运河”3个核心板

复性工作,在已有基础上对大运河线性文化遗产数字资源进行数值属性、对象属性、时空关系及事件情境等的描述,建立起适用于大运河线性文化遗产的数据模型,从而提高数据的互操作性、标准化程度和跨领域的可扩展性。本文所构建的模型共包括8类,39个属性,这些属性来源于CRM、EDM、DC、OWL、FOAF、RDFS、SKOS、GEO等,且包括10个自定义属性,如图5、图6所示。表1以实例形式展示大运河线



表 1

京杭大运河线性文化遗产实例表示

项目	元素	值
京杭大运河	相关人物(edm:hasMet)	吴王夫差;秦始皇;汉顺帝;曹操;孙权;隋炀帝……
	位置信息(edm:currentLocation)	地跨北京、天津、河北、山东、河南、安徽、江苏、浙江8个省、直辖市
	始建于(edm:occuredAt)BFQ 运河构成(crm:iscomposedof)	公元前486年 隋唐大运河、京杭大运河和浙东大运河
河道	描述信息(dc:description)	全长2700千米,跨越地球10多个纬度,纵贯在中国华北平原上,通达海河、黄河、淮河、长江、钱塘江五大水系,是中国古代南北交通的大动脉,至2020年大运河历史延续已2500余年
	名字(skos:prefLabel)	通济渠(汴河、汴水)
	相关人物(edm:hasMet)	梁惠王;隋炀帝杨广
河道	发源地(edm:currentLocation)	河南省郑州市荥阳板渚
	始建于(edm:occuredAt)	隋朝605年
	相关事件(edm:wasPresentAt)	隋炀帝命令开凿运河
河道	相关内容(edm:isRelatedTo)	《虞城县志》《资治通鉴·隋纪四》《大业杂记》
	建造河道的设计师、工人等(crm:carriedoutby)	宇文恺
	描述信息(dc:description)	通济渠自河南省郑州市荥阳的板渚(今汜水镇)出黄河,经鸿沟、蒗荡渠、睢水沟通了江苏盱眙境内的淮河,全长650千米,共流经3省6市,是古代中国劳动人民创造的一项伟大工程
河道	描述河道所在地的可视化信息(crm:represents)	
	目前管理机构(crm:hasCurrentKeeper)	河南省文物局;安徽省文物局;江苏省文物局
	通向(*ca:leadsTo)	淮河
河道	起点(*ca:startPoint)	河南省郑州市荥阳的板渚
	终点(*ca:endPoint)	江苏盱眙
	途经(*ca:passThrough)	荥阳、郑州市、中牟县、开封市、杞县、睢县、宁陵、商丘市、虞城、夏邑、永城、濉溪、宿州、灵璧、泗县、泗洪、盱眙
河道	主要支流(*ca:mainTributaries)	鸿沟、蒗荡渠、睢水
	名字(skos:prefLabel)	苏州盘门(蟠门)
	相关人物(edm:hasMet)	吴王阖闾
文物	位置信息(edm:currentLocation)	江苏省苏州市姑苏区东大街49号
	始建于(edm:occuredAt)	元朝1351年
	相关事件(edm:wasPresentAt)	1963年,盘门被列为苏州市文物保护单位
文物	相关内容(edm:isRelatedTo)	《吴地记》
	类型(edm:hasType)	防御守卫建筑
	保护级别(*ca:protectionLevel)	4A级景区
非遗	名字(skos:prefLabel)	锡剧
	相关人物(edm:hasMet)	传承人
	位置信息(edm:currentLocation)	无锡、常州一带
非遗	起源于(edm:occuredAt)	清乾隆、嘉庆年间
	相关事件(edm:wasPresentAt)	2008年6月7日,锡剧经中华人民共和国国务院批准列入第二批国家级非物质文化遗产名录,编号IV-103
	类型(edm:hasType)	传统戏剧

性文化遗产知识表示模型。

4.4 京杭大运河线性文化遗产语义增强

4.4.1 京杭大运河线性文化遗产结构化语义增强

1) 基于本体扩展的语义增强

本体扩展是在原有本体基础上添加新的实体、属性或关系,增加原有知识表示模型的关系丰富度、属性丰富度,提高模型数据内容表达的外延能力,以适应新的应用场景。主要方法包括基于知识体系的本体扩展和基于推理技术的本体扩展。

基于知识体系的本体扩展。本文在3.1线性文化遗产知识体系基础上,结合《中国古代建筑分类标准》《文物保护单位建筑物分类和代码》《中华人民共和国文物保护法》等多个标准规范,对京杭大运河线性文化遗产知识表示模型进行内容扩展。根据模型中核心类和情境类对象特征,进一步划分子类,如图7所示,如文物类可以分为可移动文物和不可移动文物,前者包括实物、艺术品等,后者包括水利工程遗址、古墓葬及古建筑等;事件类可以划分为展览、会议及节日事件等。其中情境类可以参考成熟的本体模型,如Event、ORG、FOAF、TIME、GEO等。大运河属性值扩

展主要基于核心类资源属性特征及资源的实际具体应用情况,并参考相关本体模型的属性加以扩展,如引入Event本体中的isSubEventOf、isSuperEventOf属性;Time本体中intervalMeets、intervalOverlaps、before、after属性;Geo本体中timezone、adjacent属性等。

在京杭大运河知识表示模型构建过程中,只将古建筑作为文物大类进行子类划分,在实际的大运河领域本体应用过程中,必然要细化资源类和属性特征,以满足用户实际需求。本文以大运河苏州盘门古建筑为例,阐释基于知识体系的本体扩展过程。依据建筑叙事学理论(Architectural Narratives),借鉴陆邵明^[24]学者提出的建筑叙事学本体建构维度,将古建筑划分为空间叙事主体(Agent)、空间叙事载体(Carrier)、空间叙事结构(Structure)、空间叙事风格(Style)4个类。叙事主体(Agent)又可以划分为人物(Person)和组织(Organization)两个子类;空间叙事载体(Carrier)分为虚拟空间、空间语言及空间实践3个子类;空间叙事结构(Structure)包括金字塔式、迷宫式、镜像式、迭代式、离散式以及线性模式等概念类别,用skos:Concept表示;空间叙事风格(Style)包括历

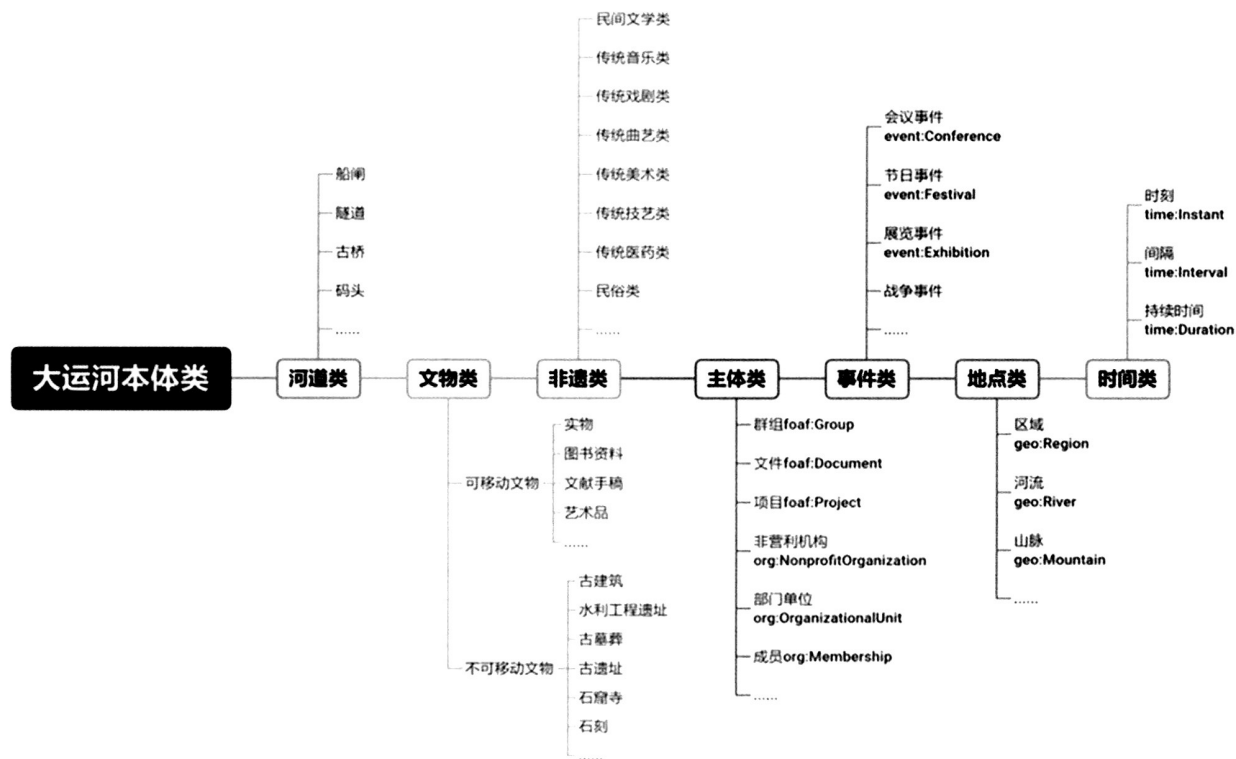


图7 京杭大运河本体类扩展

史图式风格、自然图式风格、现实图式风格、乌托邦式风格等,用 skos:Concept 表示,如图 8 所示。对于属性扩展可以根据实际具体情况进行自定义扩展,或参考本体词表进行扩充。

基于推理技术的本体扩展。推理是对本体中概念及关系进行逻辑推理和判断,自动生成新概念及关系,以此扩展本体语义表示能力。通常采用基于规则和语义相似度的本体方法,根据规则和实体间语义相似度推断出新概念、实例属性及关系。表 2 展示了两组基于规则的本体推理描述性语言。在实际应用过程中,可以利用 OWL-RL 推理引擎定义复杂的一些规则来推导新知识。

基于语义相似度的本体推理主要应用于本体的

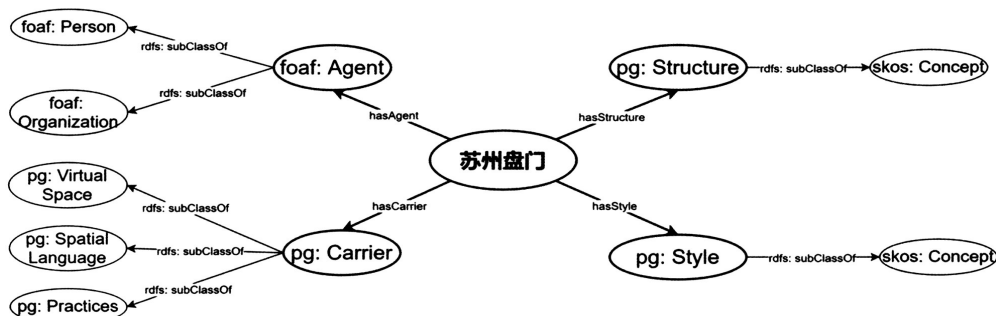


图8 苏州盘门古建筑类扩展

表 2

基于规则的本体推理

RDF 规则语言表述	规则(RDF 三元组)	推导后(三元组)
IF?xadjacent?y ?ybelong_to?z THEN?xbelong_to?z	〈运河,相邻,杭州〉〈杭州,属于,浙江省〉	〈运河,属于,浙江省〉
IF?xin_canal?y ?yhas_building?z ?ztype_istemple THEN?zis_on_canal?x	〈杭州,在,运河沿线〉〈运河沿线,有古建筑,灵隐寺〉 〈灵隐寺,类型,寺庙〉	〈灵隐寺,在,杭州〉

实例拓展,将外部数据源引入到本体知识框架体系中,丰富本体知识内容。以大运河古建筑“苏州盘门”为例,如图 9 所示,在大运河古建筑本体基础上,采集有关苏州盘门游记、历史书籍及网络资源等,用自然语言处理技术进行数据预处理,提取简介、建筑风格、特征等数据,再利用 LDA 或 LSA 主题建模抽取文本内容主题类别,比如有建筑风格、历史沿革、旅游信息及地理环境等主题,用余弦相似度或欧几里得距离计算该主题类别与大运河古建筑本体中的实体相似度,降序排列,选取最优实体概念扩展原有本体。

2)基于实体链接的语义增强

实体链接是将文本实体链接到知识库对应实体的过程,是自然语言处理和信息检索等领域的重要

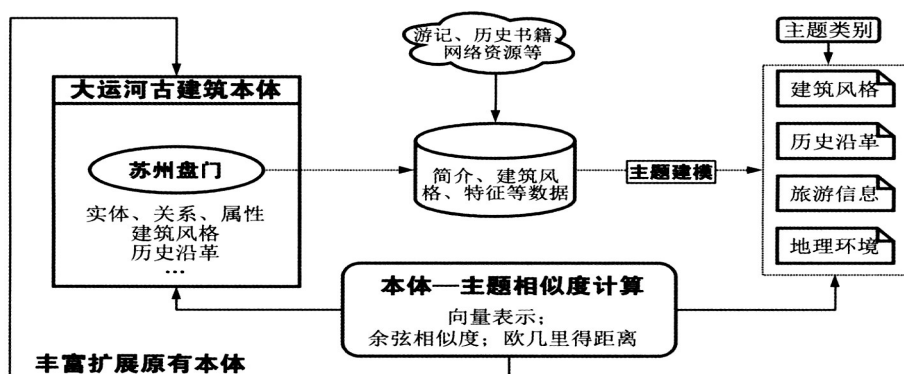


图9 基于本体-主题相似度方法的本体扩展

任务。在大运河线性文化遗产本体模型基础上,以“苏州盘门”为例,通过链接 VIAF、TGN、ASCDC、AAT、LOCAH 等外部词表来增强盘门实体语义内涵,如图 10 所示。通过 OWL 中的 owl:sameAs 属性,将本地实体“盘门”与 DBPedia、WikiPedia、Baidu Baike 等外部数据来源进行实体关联,通过 URI 链接到本地,如图 11 所示。

4.4.2 京杭大运河线性文化遗产情境化语义增强

1) 基于时间序列的语义增强

时间维度是历史事件叙事过程中不可或缺的要

素,助力学者厘清事件演变过程中的时序变化特征,深化对事件来龙去脉的认知。大运河历史悠久,具有典型的时间流动性特征。本文以大运河(江苏段)河道演变事件为轴心,融合多模态数据,如文本、图像、视频等资源,从大尺度和长时段视角揭示大运河河道的历史演变特征,并将大运河河道知识表示模型映射到河道演变事件的时序建模过程中,采用 TimeLineJS 工具可视化呈现大运河线性文化遗产河道演变的时序特征,如图 12 所示,数据源

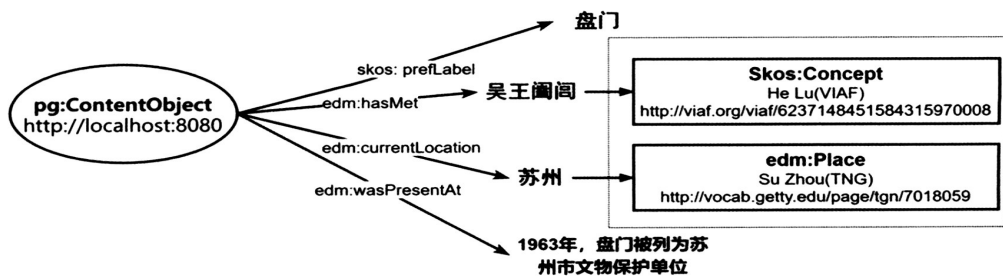


图 10 链接外部词表

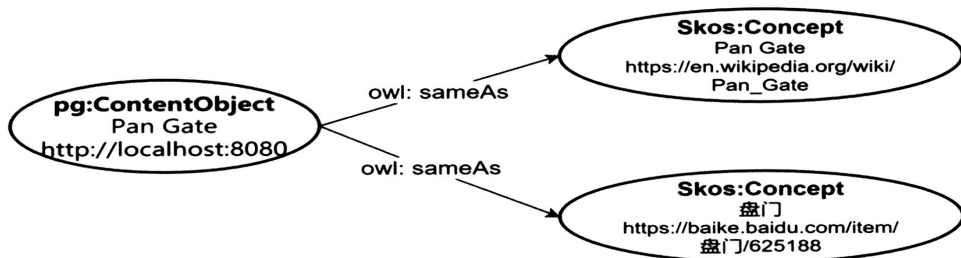
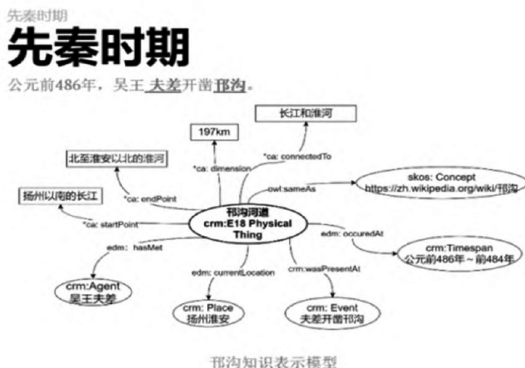


图 11 链接外部知识库



存档扩展视频资源链接: www.bilibili.com/video/BV1c741...
资料来源: 国家大运河公园小程序



邗沟知识表示模型



图 12 TimeLineJS 可视化大运河江苏段河道演变(先秦时期)

云平台相关资料,同时增加外部视频链接以扩展河道知识内容。上部右侧展示的是河道开凿事件的本体知识表示模型,以“邳沟河道”为例,结合本文构建的大运河知识表示模型,构建“邳沟河道”知识网络

表3 TimeLine.IS数据源

Year	End Year	Headline	Event Text	Media
-486	-221	先秦时期	公元前486年,吴王夫差开凿邗沟	
-220	589	秦汉南北朝时期	秦始皇时期建成丹徒水道; 西汉吴王刘濞开茱萸沟; 汉武帝年间接通苏嘉运河; 东汉广陵太守陈登改凿西道; 东晋形成“仪真运河”; 西晋练湖形成	
581	907	隋唐时期	隋炀帝下令开凿隋唐大运河	
960	1128	宋时期	宋天禧开扬州城南运河; 宋庆历中自淮阴开新河至洪泽镇; 宋元丰开龟山运河; 江南运河于宋大圣在京口闸之东建小京口; 宋元祐锡澄运河初具雏形; 宋绍熙开德胜河	
1271	1644	元明时期	元代开通京杭运河,济州河,会通河; 明永乐疏浚仪真、瓜洲运河河道; 明隆庆开南阳新河; 明万历于埕城东开辟新渠,开洲河、草湾河、永济新河、界首月河	
1662	1912	清(康熙年间后)时期	清康熙开皂河40里连接邗河和黄河; 康熙四十二年开通中运河; 清雍正开小河港(即新孟河)	

图,对其事件要素进行结构化呈现。其他河道知识网络图如表3 Media列所示。通过揭示大运河河道历史演变的时序特征,深化对河道形态、开凿事件和重要历程的认知。通过时序建模和可视化展示,能够连贯地呈现大运河河道在不同时期的变化,使历史事件与具体时代相结合,丰富知识表示模型。这不仅帮助学者厘清历史事件的连续性,还提供了更深入的历史背景信息,可以更加准确地理解和感知大运河作为线性文化遗产的重要价值。

2)基于空间POI的语义增强

兴趣点(Point of Interest, POI)是地理信息系统领域的一个概念,通常由地理坐标和一些附加属性(如名称和类别)定义,指地理空间中的特定位置场所,

被用作地图上的标记点,作为地理信息的重要组成部分,已成为人们日常生活中获取信息和进行位置服务的重要途径。本文通过POI数据来增强文化遗产数据中所包含的空间语义内涵,使POI更具有可读性、可理解性和可操作性。

京杭大运河通常被划分为通惠河与北运河段、南运河段、梁济运河段、南四湖区段、不牢河段、中运河段、里运河段及江南运河段^[25]。本文以江南运河段为例进行空间POI遗产数据的语义增强,采用TimeMapper5工具对江南运河段的文化遗产进行空间POI数据呈现,如图13所示,并以苏州盘门古建筑为例,进行POI数据的语义增强,如图14所示,首先将盘门古建筑与古建筑知识表示模型进行关联,其



图13 TimeMapper可视江南运河段文化遗产空间POI位置信息

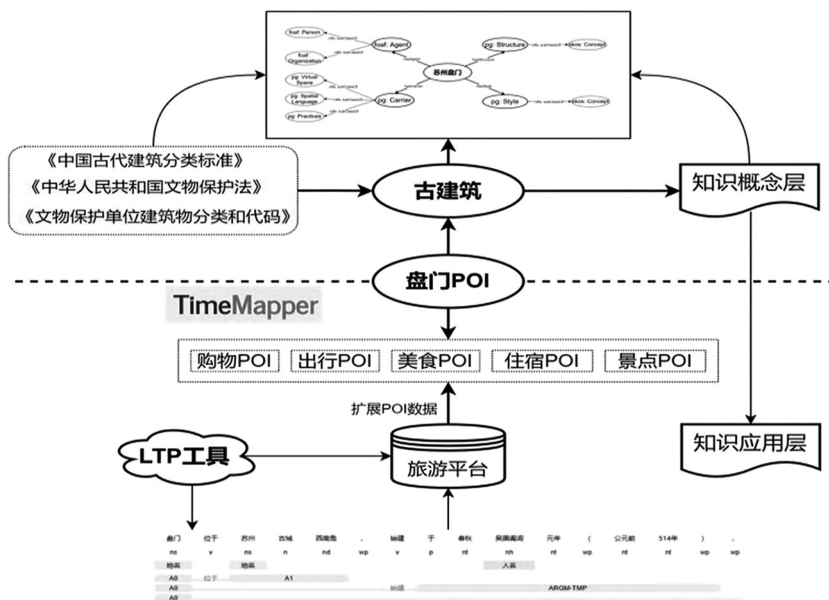


图14 苏州盘门空间POI数据语义增强模型

次增加外部实体超链接,链接百度地图盘门POI位置,利用百度地图API获取盘门固定半径之内的POI数据,包括围绕盘门景区的“衣、食、住、行、游”等维度数据。同时,收集外部旅游数据平台如马蜂窝、携程、同程、飞猪等网站上关于盘门的游记文本信息,采用地名实体识别工具LTP进行POI数据扩展,从而构建围绕盘门的POI数据语义增强模型,实现盘门POI数据从知识概念层到知识应用层的转换。通过空间维度的语义增强,运用POI数据进一步丰富线性文化遗产数据的空间语义内涵,使特定位置场所更具实际含义,具备更高的可读性、可理解性和可操作性,提升地理信息的交互性与实用性,使人们能够更深入地了解、感知并参与文化遗产的传承与传播,进一步拓展了地理信息的应用价值。

5 结论

随着信息技术的飞速发展和数字化时代的到来,文化大数据化已成为发展趋势。线性文化遗产作为我国丰富而独特的文化瑰宝,蕴含着丰厚的历史价值和文化内涵。本文在此背景下,以京杭大运河线性文化遗产为例,深入到线性文化遗产内容层面,深度剖析线性文化遗产知识体系的层级结构,建构线性文化遗产知识表示模型,并依据表征理论和情境认知理论,从结构化语义增强和情境化语义增强视角对线性文化遗产进行语义增强,创新性地提出“时空”双重维度语义增强路径。其中,时间维度的语义增强主要通过在大运河(江苏段)河道演变事件中引入时间维度,融合多模态数据如文本、图像、视频等,采用TimeLineJS工具的时序可视化呈现,以大尺度和长时段视角揭示大运河河道的历史演变特征,有助于更清晰地厘清不同历史时期的河道开凿事件,使事件的时序变化特征得以凸显,为学者提供更具体、更完整的历史叙事。空间维度的语义增强主要通过POI数据实现,利用TimeMapper工具对文化遗产的空间POI数据进行可视化呈现,以及通过盘门古建筑实例进行语义增强,通过将古建筑与知识表示模型关联、超链接至百度地图、整合周边维度数据等方式,从多个维度丰富POI信息,旨在使POI数据更富有可读性、可理解性和可操作性,进一步丰富地理空间中的特定位置场所的信息,为人们提供

更丰富的位置服务和文化遗产的互动体验,加深对文化遗产的认知与了解。在后续的研究中,笔者将会继续完善本文所建构的线性文化遗产知识体系,通过量化分析和统计分析方法对线性文化遗产POI数据进一步挖掘和处理,从实证角度探索更多的可能性。

参考文献:

- [1]单霁翔.大型线性文化遗产保护初论:突破与压力[J].南方文物,2006,(3):2-5.
- [2]Gonçalves A R, Dorsch L L P, Figueiredo M. Digital Tourism: An Alternative View on Cultural Intangible Heritage and Sustainability in Tavira, Portugal[J]. Sustainability, 2022, 14(5): 2912.
- [3]Brinia V, Psoni P, Ntansiou E K. How to Instill Cultural Values in the New Generation Through Cultural Promenades and Ancient Drama: A Field Research[J]. Sustainability, 2019, 11(6): 1758.
- [4]李飞,邹统钎.线性文化遗产重要节点旅游发展模式研究——以云南碧色寨为例[J].地理与地理信息科学,2022,38(5):138-144.
- [5]刘军民,张清源,巩岳,等.国土空间规划中线性文化遗产的保护利用研究——以咸阳市为例[J].城市发展研究,2021,28(3):7-13.
- [6]俞孔坚,奚雪松,李迪华,等.中国国家线性文化遗产网络构建[J].人文地理,2009,24(3):11-16,116.
- [7]刘美杏,徐芳.古道文化遗产信息资源元数据标准制定——以萧贺古道为例[J].情报资料工作,2019,40(4):77-83.
- [8]孙绍丹,邓君,常严予,等.近代报纸资源细粒度语义描述模型设计及应用——以《盛京时报》为例[J].图书情报工作,2022,66(7):35-46.
- [9]李章超,何琳.文化遗产语义组织研究进展[J].图书情报工作,2020,64(7):4-12.
- [10]侯西龙,王晓光,翁梦娟.文化遗产领域知识纲要研究[J].图书情报工作,2022,66(22):106-118.
- [11]孙绍丹.数字人文视域下历史报纸资源语义化知识组织研究[D].长春:吉林大学,2023.
- [12]Wang X, Chang W, Tan X. Representing and Linking Dunhuang Cultural Heritage Information Resources Using Knowledge Graph[J]. Knowledge Organization, 2021, 47(7): 604-615.
- [13]宋雪雁,张伟民,张祥青.基于档案文献的清代祭祀礼器知识图谱构建研究[J].图书情报工作,2022,66(3):140-151.

[14]张卫,王昊,李跃艳,等.面向非遗文本的知识组织模式及人文图谱构建研究[J].情报资料工作,2021,42(6):91-101.

[15]曾蕾,谭旭.数据的语义增强——解读图档博支持数字人文的新动向[J].数字人文研究,2021,1(1):65-86.

[16]Damjanovic-Behrendt V, Kurz T, Westenthaler R, et al. Semantic Enrichment: The Key to Massive and Heterogeneous Data Pools[C]//Proceeding of the 20th International IEEE ERK (Electro-technical and Computer Science)Conference, 2011: 413-416.

[17]Abgaz Y, Souza R R, Methuku J, et al. A Methodology for Semantic Enrichment of Cultural Heritage Images Using Artificial Intelligence Technologies[J]. Journal of Imaging, 2021, 7(8): 121.

[18]Tan X, Luo X, Wang X, et al. Representation and Display of Digital Images of Cultural Heritage: A Semantic Enrichment Approach[J]. Knowledge Organization, 2021, 48(3): 231-247.

[19]World Heritage List[EB/OL]. <https://whc.unesco.org/en/list/?&>, 2023-05-04.

[20]张书颖,刘家明,朱鹤,等.线性文化遗产的特征及其对旅游利用模式的影响——基于《世界遗产名录》的统计分析[J].中国生态旅游,2021,11(2):203-216.

[21]李飞.线性文化遗产空间结构演化研究——兼述旅游于其中的影响[J].地理与地理信息科学,2019,35(5):133-140.

[22]俞肇元,袁林旺,吴明光,等.地理学视角下地理信息的分类与描述[J].地球信息科学学报,2022,24(1):17-24.

[23]孙婷,黄泰.多维视角下大运河遗产认知与保护传承[N].中国社会科学报,2023.

[24]陆邵明.当代建筑叙事学的本体建构——叙事视野下的空间特征、方法及其对创新教育的启示[J].建筑学报,2010, (4):1-7.

[25]俞孔坚,李迪华,李海龙,等.京杭大运河国家遗产与生态廊道[M].北京:北京大学出版社,2012.

Construction of Knowledge Representation and Semantic Enrichment Framework for Linear Cultural Heritage Resources in the Context of Cultural Big Data

Sun Shaodan Zhang Liman Zhu Xiang

Abstract: [Purpose/Significance] This study aims to construct a knowledge representation and semantic enhancement framework for linear cultural heritage resource, which provides guidance for cultural institutions to carry out intelligent data construction of cultural heritage and promotes the innovative development of excellent traditional culture. [Method/Process] Based on the distilled characteristics of "resource communalism, temporal flow, and spatial chaining" in extracting linear cultural heritage, this study analyzed the hierarchical structure of the knowledge system of linear cultural heritage. It constructed a knowledge representation model for linear cultural heritage resources. Basing on representation theory and situational cognition theory, it semantically enriches linear cultural heritage data from the perspectives of structured semantic enhancement and contextual semantic enhancement. It innovatively proposed a dual-dimensional semantic enhancement path of "time and space". Finally, took the Grand Canal linear cultural heritage as an example, conducted a case study on knowledge representation and semantic enhancement of the Grand Canal. [Result/Conclusion] The knowledge representation model and semantic enhancement framework constructed in this study are conducive to multidimensional semantic organization and in-depth mining of linear cultural heritage data, which can enhance the semantic connectivity and extension of linear cultural heritage contents.

Key words: Cultural big data; Linear cultural heritage; Knowledge representation; Semantic enrichment; The Beijing-Hangzhou Grand Canal