



融合大语言模型的中国古代山水志史料知识重构 与可视化路径研究

——以《续修大岳太和山志》为例

周树斌 程双玲*

摘要 山水志是中国山水文化记忆的重要载体,结合数智技术探索中国山水志史料的知识重构与可视化路径有助于进一步激活其文化价值和记忆价值。文章围绕山水志史料,设计了以“物质维度—象征维度—功能维度”为主线的知识图谱构建方案,提出了涵盖资源采集、知识表示、知识抽取及知识可视化四个模块的知识重构与可视化研究路径。在实证方面,以《续修大岳太和山志》为例,基于大语言模型与低代码平台 Huggingfists 实现知识图谱全流程构建。通过可视化呈现武当山文化,实现其数字记忆场景重构。文章在山水志史料数字化开发与山水文化传承方面提供了有益的探索路径,也为历史人文领域引入大语言模型提供了可行方案。

关键词 山水志 知识图谱 大语言模型 知识重构 数字人文

分类号 G250.7

DOI 10.16603/j.issn1002-1027.2025.05.007

引用本文格式 周树斌,程双玲.融合大语言模型的中国古代山水志史料知识重构与可视化路径研究——以《续修大岳太和山志》为例[J].大学图书馆学报,2025,43(5):54-66.

近年来,全球范围内文化数字化进程加速推进。以美国、英国、澳大利亚等为代表的西方国家率先布局,持续推进文化数字化进程,旨在通过数字技术实现文化资源的形态转化与价值重塑^[1]。在此背景下,我国立足国家文化发展的战略需求,于2022年颁布《关于推进实施国家文化数字化战略的意见》^[2],明确提出建设公共文化数字资源库群和国家文化大数据体系的战略目标,标志着文化数字化建设已正式被纳入中国国家发展战略框架。作为中国传统文化资源的重要组成部分,地方志承载着“资治、教化、存史”的功能,系统记载了特定区域自然、政治、经济、社会及文化的沿革与现状^[3]。据不完全统计,我国现存地方志8000余种,累计11万余卷^[4],丰富的地方志资源体系不仅为研究中国地域社会文化提供了第一手资料,更为构建中华民族集体记忆、传承历史文脉奠定了坚实的文献基础。山

水志作为地方志的特殊门类,聚焦于山水地理及其文化内涵的深度记述。相较于综合性的府、州、县志,山水志更为详实、具体地记录了与山水相关的文化历史资料,突出展现了地方山水文化的记忆,是研究人地关系与区域文明的重要资源。然而,当前山水志研究面临双重困境:一方面,现存山水志数量仅占地方志总量的一小部分,史料来源的稀见性导致其研究基础薄弱;另一方面,学界长期将其定位为辅助性史料,缺乏本体论层面的系统性研究^[5]。这种研究现状与山水志在数智时代的潜在价值形成显著反差,亟待学界予以重视。

随着信息技术的迭代演进,以人工智能、数字孪生、区块链和云计算为代表的新一代数字技术,正在引发人文社会科学研究范式与方法论体系的深刻变革。在此背景下,数字人文作为计算机技术与人文研究交叉融合的新兴学术范式^[6],不仅拓宽了传统

* 内蒙古自治区社会科学基金项目“《中华民族共同体概论》文化遗产多模态知识组织与知识聚合研究”(编号:2025PTWY13)的研究成果之一。

通讯作者:程双玲,邮箱:chengsl20@naver.com。



人文学科的研究疆域,更为文化资源的数字化保护与创新传承提供了全新的技术路径。在知识组织领域,知识图谱技术凭借其结构化知识表征能力,已成为数字人文研究的重要支撑^[7]。大语言模型(Large Language Model, LLM)通过自监督学习机制,能够深度解析文本中隐含的概念体系与语义关联特征,在语义识别和信息抽取方面表现出显著优势。因此,将大语言模型与知识图谱技术相结合,可以实现更准确、更高效、更深层次的知识图谱构建,为数字人文研究提供更加可靠的知识处理方法。

文章以山水志作为研究对象,结合知识图谱的知识表示与描述能力,融合大语言模型进行知识抽取,深度揭示山水志史料中的实体和实体间的关系,以可视化的方式呈现史料资源的内部知识结构,实现知识重组和场景再现。同时,结合“记忆之场”(Sites of Memory)理论,进一步深化对山水志中蕴含的文化记忆和自然景观的解构与重构。在实证层面,文章以《续修大岳太和山志》^[8]为具体案例,结合大语言模型和提示语工程,通过低代码平台 Huggingfists 实现知识图谱的全流程构建。与传统的人工标注和数据处理方法相比,低代码平台显著提升了效率,并有效降低了技术门槛,使更多研究人员能够便捷地参与到古籍数字化与智能处理工作中。随后,结合知识发现和可视化分析,场景化再现了武当山的历史文化记忆。这种跨学科方法不仅为山水志的数字化转化提供了新的研究框架,也为传统古籍的智能处理和文化遗产的数字化保护提供了可复用的研究范式。通过结合大语言模型与低代码平台,文章的研究路线亦展现了对其他类型地方志及文化遗产资源数字化转化的广泛适用性,可为同类研究提供有价值的参考。

1 相关研究

数字人文起源于人文计算,是以人文数据为基础,以计算机技术为支撑,辅助研究者开展人文研究而形成的新型跨学科研究领域^[9]。随着数字人文研究的不断深入,采用本体、关联数据、知识图谱等语义网技术,对古籍、档案、非物质文化遗产等进行语义表示和组织已成为新的研究范式。知识图谱能够实现数据的语义化组织、可视化呈现和智能化分析,以直观、清晰的方式展现复杂的知识结构^[10]。现有研究中,张云中等构建了井冈山革命知识图谱,并提

出了基于知识图谱的井冈山革命文化资源数字叙事模型^[11];谢玮和杨家瑶构建了耕织图像知识图谱,从多个角度展示耕织图像所涉及的知识内容^[12]。此外,知识图谱技术也广泛用于历史名人^[13-14]、古代战争^[15]、文化遗产^[16]等人文领域的研究。数字人文的发展,也为地方志的知识挖掘和组织不断引入新的方法和研究视角。针对地方志资源知识价值利用率极为有限的现状,熊欣等探讨了面向方志知识图谱的自动化术语抽取^[17];部分研究则聚焦于地方志中的特定领域,如围绕物产资源展开语义化知识组织^[18]和知识库系统构建^[19]。此外,也有学者将知识图谱与关联数据等技术融合应用于方志资源的语义挖掘。例如,高劲松等对山水志史料资源进行语义知识关联^[5];徐晨飞对方志书目提要进行了深度语义化知识组织^[20]。然而,这些研究普遍依赖人工标注,难以满足大规模史料建模与语义理解的需求。

随着人工智能的兴起,大语言模型凭借其强大的学习和推理能力,为知识处理和抽取提供了新的技术路径。张兴旺等提出了基于大语言模型的古籍文献知识组织框架,并验证了其在知识表达、知识推理与智能服务方面的应用可行性^[21];陈和利用大语言模型探索了参考文献自动识别与著录信息抽取,指出大语言模型具有低使用门槛、较高的识别准确率以及抽取效率等优点^[22]。鉴于知识图谱与大语言模型在语义表示和分析方面的互补优势^[23],越来越多的学者将两者结合进行研究。例如,孙翌和刘音融合知识图谱与大语言模型,构建了面向口述历史档案资源的问答系统^[24];周正达等则基于大语言模型和提示工程,构建了非遗陶瓷工艺知识图谱并进行了应用场景探索^[25]。

尽管现有研究已在知识图谱构建中探索大语言模型的应用,并在术语识别与实体关系抽取等环节展现出优势,但在面向人文文本的深层语义建模方面仍显不足。尤其对于承载历史记忆与文化象征的地方志文本,现有方法往往侧重于显性信息的结构化处理,对其复杂语义网络的系统挖掘与表达仍显欠缺。山水志作为地方志的重要组成部分,系统记录了自然景观、社会经济、文化生活等各方面的内容,涵盖了草木、山石、沟壑、寺庙、道观、河川、湖泊等自然与人文要素^[26-27]。作为对山水文化的直接记录与重要记忆载体,山水志不仅具有较高的文献价值,也承载着



深厚的文化记忆。随着文化符号和社会功能的不断叠加,山水不仅作为自然景观存在,更是具有象征意义和功能的“记忆之场”。法国历史学家皮埃尔·诺拉(Pierre Nora)提出,记忆之场是“一切在物质或精神层面具有重大意义的统一体,经由人的意志或岁月的力量,这些统一体已经转变为任意共同体的记忆遗产的一个象征性元素”^[28]。作为文化记忆的承载体,记忆之场具有物质性、象征性、功能性三种特性^[29],既扎根于具体存在,又承载特定的社会意义和文化价值。因此,记忆之场理论被广泛应用于古迹等场所的记忆重构^[30]、文化遗产的保护与更新^[31]、档案史料的解构与重构^[32]等领域,为在山水志研究中引入该理论提供了重要的参考。

综上所述,现有研究探索了在地方志研究中应用数字人文方法的可行性,但方法主要依赖既定的知识图谱构建流程,并需要大量人工标注,对文本深层的文化记忆与象征意义关注不足。基于此,文章引入记忆之场理论,结合大语言模型技术,构建更为高效且系统的知识图谱,探索山水志史料的知识重构与文化记忆场景再现路径,为山水志研究提供新的思路。

2 研究框架设计

山水志作为记录山水文化的重要载体,其文化价值和记忆功能高度契合诺拉的记忆之场理论。中

国传统山水文化具有“形象意”三维特征^[33],文章认为该特征可与记忆之场的物质性、象征性和功能性相对应。其中,“形”体现山水志的物质性,以自然景观、建筑遗存等物理实体为载体,承载并呈现文化记忆;“象”反映了山水文化的象征性,通过对景观、人物、事件的语义关联,揭示了文化符号与历史叙事之间的深层联系;“意”则展现了山水文化的功能性,通过记忆场景的动态再现与传播实现了文化精神的传承与延续。在中国古典文论和美学中,“意”通常指艺术作品所蕴含的主观情感、精神境界和深层意蕴。文章在山水文化语境中对其加以拓展,使“意”不再局限于静态、孤立的精神境界,而是涵盖表达、沟通、教化与传承等动态功能。这一理解与诺拉“记忆之场”理论中的“功能性”维度高度契合,即记忆场域如何被建构、利用与诠释,从而实现其社会文化功能。因此,山水志知识图谱的构建实际上是从山水文化史料资源的解构到记忆场景重塑的过程,从物质维度提供史料资源的数字化依据,以象征维度实现知识单元的语义化关联,以功能维度完成知识的可视化呈现与语义发现。基于记忆之场理论与山水文化的适配性,文章设计了以“物质维度—象征维度—功能维度”为语义主线的知识图谱构建方案,并提出了资源采集、知识表示、知识抽取和知识可视化的四模块知识重构路径,具体如图1所示。

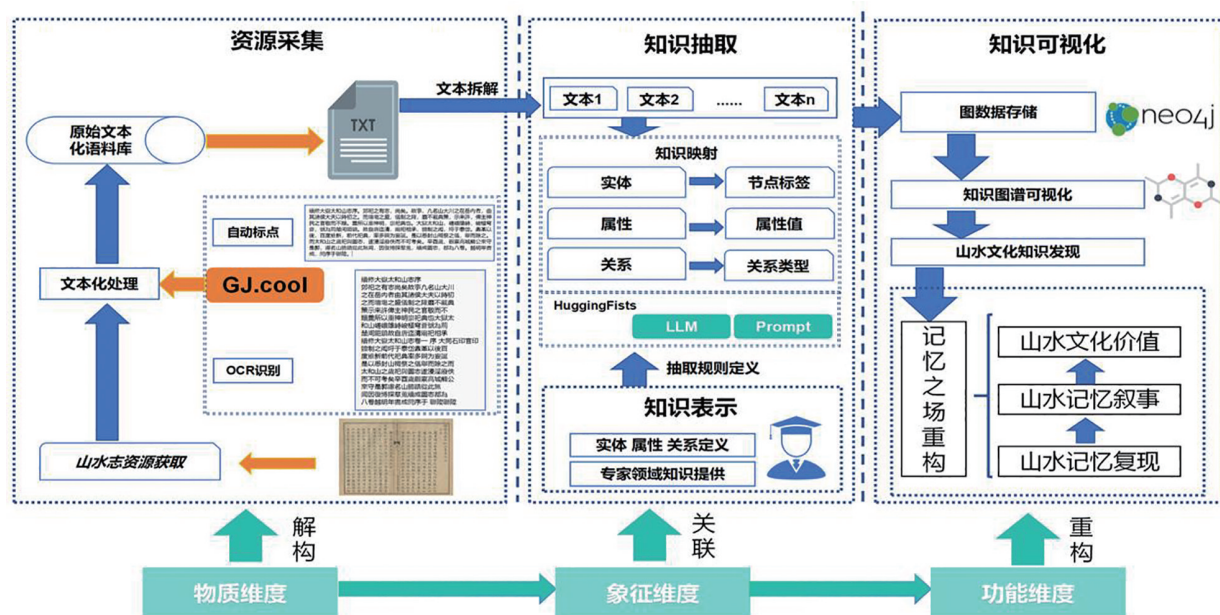


图1 融合大语言模型的山水志知识重构与可视化框架



2.1 物质维度:资源采集

在山水文化记忆的建构与传承中,“形”作为一切文化意义的原始物质载体,构成了记忆之场理论中“物质性”的基石。中国古代山水志不仅是记录山水文化的文本,其本身(如古籍的纸张、墨迹、版式)以及其所描绘的山川、石刻、建筑遗存等实体景观,共同构成了山水文化记忆最根本的物质依托。因此,物质维度的资源采集主要是对这些物质载体进行高精度数字化映射的过程,该过程的关键在于将纸本山水志书进行文本化处理。文本化处理的第一步是利用光学字符识别(OCR)技术将史料资源进行文本化处理得到原始文本数据。而后是文本预处理,主要是面向原始文本进行标点、分词等工作,得到规范化、机器可读的原始文本语料库,成为后续山水文化史料知识要素提取的母本语料。文章中,OCR以及自动标点功能主要通过“古籍酷”(GJ. cool)平台实现。

2.2 象征维度:知识表示与知识抽取

从象征维度来看,山水志中的“象”并非孤立的客观记录,而是被赋予了深厚的文化意义和情感价值的符号系统。它们通过特定的叙事、比喻和分类方式,将散乱的物质实体(“形”)组织成表达特定世界观和价值观的有序结构。因此,象征维度的知识化过程,其核心在于解读和重构这些历史叙事与文化符号的语义关联,揭示其如何承载和传递集体记忆。知识表示和知识抽取是这一过程的基础。知识表示是通过明确山水志文本核心实体、属性和关系,完成对领域内知识概念的规范化描述和完整性揭示,从而为知识抽取提供模式规范。知识抽取是指利用自然语言处理和机器学习技术,对文本数据进行实体识别、属性和关系抽取。通过识别文本中的实体,如景观地名、历史事件、人物等,以及它们之间的关系,可以将史料资源中的重要要素提取出来,并形成结构化数据。现有研究在这一过程中需要注意对实体、关系和属性数据进行标注和质量控制,确保数据的准确性和一致性,具体可通过人工标注、自动验证和数据清洗等方法来实现,标注和质量控制过程则可以借助专家知识和规范进行。该过程较为繁琐复杂,不仅需要大量人工参与知识标注工作,还依赖大规模的文本训练任务,且难以充分挖掘领域知识的深层内涵。

近年来,大语言模型以其强大的自然语言处理

能力和对话生成能力,为完善知识抽取工作带来了新的可能性。大语言模型能够自动抽取知识、生成文本,并以对话的方式与用户进行交互。通过合理设计的提示语,大语言模型可以完成复杂实体、属性、关系的抽取。考虑到大语言模型的上述优势,文章基于低代码平台 Huggingfists,通过设计提示语进行山水志文本的知识抽取,根据知识表示的模式化约定,识别出文本数据集中的实体、属性和关系。最后,对抽取得到的离散知识进行融合,并通过知识表示的规范化描述,实现对实体、属性和关系的语义化关联。实体概念和实例对应于图数据库中的节点标签(Node Labels),属性对应于属性值(Property Keys),关系对应于关系类型(Relationship Types),基于上述知识映射可完成知识的序化聚合形成知识三元组。

2.3 功能维度:知识可视化与知识发现

从功能维度来看,山水志知识图谱的构建旨在实现文化记忆的动态再现、语义化检索与深度知识发现,其核心是通过技术手段激活并外化“意”所蕴含的精神功能,实现文化价值的延续与传播。本研究将抽取并结构化的山水文化知识(实体、关系、属性)以三元组形式存储于 Neo4j 中,为其后的交互式可视化、语义关联探索与深层文化分析奠定了数据基础。完成知识存储后,Neo4j 的图数据库可视化功能被应用于展示山水文化的知识图谱,从功能维度实现了知识可视化与知识发现的核心目标。

知识可视化的关键是通过实体“点”和关系“线”的深度链接,从“面”的角度整合散落的历史记忆片段,形成全景式的文化语义网络。例如,通过知识图谱,能够直观展示地方山水的核心景观与其相关的历史事件以及文化符号的关联路径。这种多视角、多层次的可视化,不仅使山水志知识变得更加直观,也为文化研究提供了完整的语义上下文。从“场”的角度来看,知识图谱的交互式可视化功能能够将“点一线一面”深度融合,动态再现山水文化的记忆场景。这种动态可视化不仅实现了物质维度(如景观与地理位置)的有序化,还通过象征维度(如文化符号)的外显化与功能维度(如历史场景重现)的综合化,形成一个多维度、多时间尺度的文化记忆空间。

Neo4j 的查询功能使研究者能够快速定位山水文化知识图谱中的隐性语义关联,如可以通过分析历史人物与相关景观之间的关系,发现人物在历史



文化传播中的作用。知识发现还能够帮助研究者追踪文化事件的时间和空间轨迹。例如,通过分析文化活动的路径,可以揭示某一类文化现象从单一山水向更广泛区域扩展的传播规律,为文化研究提供新的视角。

文章以武当山山水志的知识图谱构建为案例,结合 Neo4j 的可视化与知识发现功能,从功能维度探索了山水文化记忆场景的数字化重现与语义化分析。这不仅展示了记忆之场理论的实际应用价值,还为山水文化的数字化保护与传承提供了新的研究路径。

3 实证研究

3.1 案例数据采集与处理

武当山位于古均州(今湖北省丹江口市)境内,又名太和山、仙室山,是闻名遐迩的道教圣地,素有“亘古无双胜境,天下第一仙山”的美誉。得天独厚的地理位置和清幽缥缈的自然环境,使得武当山自古便成为修仙学道者心仪神往的世外桃源,人文与山水的交融孕育出了武当山别具一格的山水文化。武当山作为名山胜地,很早就有了修志活动,虽然修志的肇始年代已难以确考,但修志传统源远流长,成就斐然。

文章选择熊宾修等纂《续修大岳太和山志》作为山水志母体文献数据来源进行具体的实证研究,该志为民国时期的第一部武当山志,它是以清代王概《大岳太和山纪略》^[34]为底本续修而成,有承前启后之特点。该志纂成于民国十一年(1922年),同年由襄阳大同石印馆刊印,约十五万字。从内容上来看,其体例格式较为规范,全书正文共八卷,类为十二类目,每一类目皆前缀小序,后附按语,以明其旨。其中卷一为星野、图考,卷二为山川,卷三为圣纪、宫殿、祀典,卷四为仙真、物产、金石、教规、拾遗,卷五至卷八悉为艺文。卷八末尾还补录有未经具体分类的艺文若干篇。文章所选版本为“中华古籍资源库”内《续修大岳太和山志》民国十一年石印本。通过 GJ.cool 平台对其进行 OCR 处理以及标点化处理完成文本化,形成文本数据集,以作为后续研究的数据基础。

3.2 山水志史料知识表示

知识表示是在明确领域内知识概念内涵的基础上,针对领域内的实体、属性、关系进行界定。为保证知识表示的泛用性,文章通过调研《中华山水志丛

刊》等权威文献,归纳总结中国古代山水志的史料知识要素,将山水志知识实体定义为景观、人物、事件、艺文、引书、时间、空间七类,各实体的属性如表 1 所示,实体之间的关系和属性如表 2 所示。

(1)景观。景观既是山水志史料资源中的核心,同时也是文化和自然美学的融合体,其包括了峰、崖、涧、洞、水、河等山川自然景观,以及台、亭、楼、阁、桥、石刻等建筑人文景观,是山水志中山水文化的具体化、空间化表征。

(2)人物。人物是历史的创造者、传承者和记录者,通过编纂、创作和参与,为后人留下了宝贵的历史见证和文化遗产。人物与景观、艺文、事件紧密相连,他们的行为和创作丰富了山水志的内容,使之成为研究历史、文化和社会发展的重要资料。

(3)事件。事件是山水志中记录的重要历史时刻和发展过程。它们可以是自然事件,如洪水、地震,也可以是社会事件,如战争、政治变革。事件与人物、时间和空间紧密相关,它们在特定的时间和地点发生,对历史进程产生影响。

(4)艺文。艺文是山水文化传承的重要形式,它包括赋、铭、文、勅、歌、诗、书、谕、赞、诏、旨、跋、序、记等多种形式。艺文作为人物与景观的桥梁,不仅展现了作者的艺术才华和审美情趣,也反映了特定历史时期的社会风貌和文化特色,因而成为山水文化超越自然山水本身的重要载体和媒介。

(5)引书。引书是指在山水志中引用的其他文献资料,这些资料为山水志提供了历史背景、文化解释和知识补充。通过引用经典文献、历史记载和其他相关资料,山水志的编纂者能够丰富内容,增加说服力和权威性。

(6)时间。时间是描述史料资源的重要维度,包括时间点和时间段。时间点记录了具体的历史时刻,如特定年份或朝代。时间段则描述了历史的发展过程,如一个朝代的兴衰或一个时期的文化变迁。通过对时间的精确记录和描述,能更好地理解历史事件的脉络和变迁。

(7)空间。空间是描述史料资源的另一个重要维度,它包括地区和地点。地区描述了史料资源的广泛地理分布,而地点则具体到特定的地理位置。地点类是景观类的父类,因为景观往往与特定的地点相关联。通过对空间的详细描述,可以更好地理解景观的地理特征和文化背景。



表 1 实体及属性定义

实体	属性	示例
Landscape(景观)	Name(景观名称)	天柱峯
	Other Names(景观别名)	金頂, 大頂
	Title(景观类型)	峯
Person(人物)	Name(姓名)	熊賓
	Other Names(人物别名)	—
	Title(人物职官)	襄陽道道尹
Event(事件)	Event Name(事件名称)	尹喜隱居老君洞
	Event Description(事件描述)	因尹喜曾隱武當,老君過訪
	Type(事件类型)	傳說
Art Literature (艺文)	Title(艺文题名)	南巖宮
	Content(正文)	古木懸孤嶂,虛巖綴百靈。亭陰侵座爽,林影上衣青。 地寂年為日,雲屯晝欲暝。何方迴白首,於此問參苓。
	Type(艺文类型)	詩 五言律
Cited Book(引书)	Cited Book Name(引书名)	周禮
	Content(引文内容)	周禮職方氏但稱九州之鎮山,不稱嶽
Time(时间)	Chinese Historical Chronology(中国历史纪年)	萬曆八年
	Dynasty(朝代)	明朝
Place(空间)	Place Name(地名)	武當山

表 2 实体间的关系界定

关系	介绍	头实体	尾实体
Landscape and Landscape (景观与景观)	用于表述景观与景观之间的关系,如方位关系	Landscape	Landscape
Person and Landscape (人物与景观)	用于表述人物与景观之间的关系,如到访、居住等关系	Person	Landscape
Event and Landscape (事件与景观)	用于表述事件与景观之间的关系,如事件发生地关系	Event	Landscape
Event and Person (事件与人物)	用于表述事件与人物关系,如事件到人物具有参与者关系	Event	Person
Event and Time (事件与时间)	用于表述事件与时间关系,如事件到时间具有发生于的关系	Event	Time
Art Literature and Landscape (艺文与景观)	用于表述艺文与景观之间的关系,如艺文到景观具有相关景观关系	Art Literature	Landscape
Art Literature and Person (艺文与人物)	用于表述艺文与人物的关系,如艺文到人物具有被创作、记载等关系	Art Literature	Person
Art Literature and Time (艺文与时间)	用于表述艺文与时间的关系,如艺文到时间具有创作时间的关系	Art Literature	Time
Cited Book and Landscape (引书与景观)	用于表述引书与景观之间的关系,如引书到景观具有记载关系	Cited Book	Landscape
Cited Book and Time (引书与时间)	用于表述引书与时间的关系,如引书到时间具有成书时间关系	Cited Book	Time



3.3 山水志知识抽取与图谱构建

知识图谱的构建是一个系统性工程,涵盖了从文本数据的语义化抽取到图谱结构化存储的全过程。文章通过结合低代码平台 Huggingfists 和阿里的通义千问大语言模型,实现了山水志史料资源知识抽取与图谱构建的高效流程化操作。为了提升模型在古籍文本中的适应性和优化其抽取能力,文章对模型的提示设计进行了精细化处理,并进行了信效度检验以确保其可靠性。

3.3.1 Huggingfists 平台简介

现有知识图谱构建过程通常面临海量文本处理、高质量知识抽取与一致性维护的挑战,耗时且易受人工干预的影响。大语言模型的引入为这些问题的解决带来了新的可能性,其出色的泛化能力可以高效地完成复杂语义知识的抽取。为进一步降低技术门槛,文章选用低代码平台 Huggingfists 作为知识图谱流程化构建的核心工具。基于该平台,用户可以以一致且可视化的方式构建结构化、半结构化及非结构化数据的处理与分析流程。平台集成了丰富的人工智能算子,覆盖了读取、写入、处理、分析等各类数据处理及应用需求,显著降低了工程实施成本,使得即便是非技术背景的研究者,也能够高效地完成复杂的文本处理任务,从而大大降低了技术门槛。具体而言,Huggingfists 平台通过其图形化界面和模块化设计,允许用户以最简便的方式进行数据输入、模型训练和知识图谱构建。平台的低代码特性使得用户无需深入编写复杂代码即可完成数据处理和知识抽取任务。而大语言模型通过在海量文本数据中的预训练,展现了出色的自动化知识抽取能力,尤其是在实体识别、关系抽取以及语义推理方面。与现有方法相比,模型的零样本学习能力使其能够在没有大量标注数据的情况下,也能精准地抽取山水志文本中的实体、属性和关系,为知识图谱的构建提供了更高的精度和效率。

文章使用 Huggingfists 平台结合大语言模型的算子功能,设计并优化了从文本数据读取到知识图谱构建的操作流程。此平台的应用显著提升了知识抽取的效率和广度,同时确保了系统的可扩展性和操作的简易性。

3.3.2 提示工程准备

提示工程准备是确保大语言模型准确理解和处理特定任务的关键步骤,在山水志知识图谱构建中尤为重要,其核心在于提示语的设计。文章选择阿里的通义千问模型作为大语言模型,通过其网页对话界面设置提示语,以引导模型准确抽取和理解文

本中的关键信息。在开展提示工程之前,需要对大语言模型的能力进行评估,明确其在实体识别、属性抽取、关系识别等方面的性能。

在设计提示语之前,首先要明确任务的具体目标。例如,是否要提取历史人物、地域信息或景观知识等。明确目标后,提示语的设计就可以围绕这些目标展开,例如:“请从以下文本中抽取出所有历史人物及其相关的事件与地点,并以三元组的形式输出: {“人物”:“姓名”,“事件”:“事件描述”,“地点”:“事件发生地”}。”其次,提示语的设计应尽量简洁明了,同时充分引导模型进行所需任务的处理。对于复杂任务,采用结构化的提示语可以帮助模型理解抽取的方式和格式。例如,针对“景观与人物的关系抽取”,提示语可以设计为:“请帮我分析下列文本中的景观与人物之间的关系,并根据关系方向输出三元组,如‘居住关系’时,人物在前,景观在后,关系定义为‘居住’。请以 JSON 格式输出,输出格式如下: {“src”:“人物”,“relation”:“居住”,“dst”:“景观”}。”同时,合理的参数设置对于获取最佳结果至关重要,包括最大输出长度、温度参数、Top-k 和 Top-p 采样等,以控制模型输出的创造性、随机性和多样性。如温度参数设置较高时,模型的输出可能更具创造性和多样性,但也可能导致信息不准确;温度参数设置较低时,模型的输出则较为保守,准确性较高。针对山水志文本的特定任务,可设置较低温度并严格控制采样范围,以确保结果的精确度。

提示工程设计需要通过多轮迭代和优化来提升模型的抽取效果。构建一个反馈机制十分重要,通过人工审核、错误分析、结果对比等手段,对模型输出结果进行评估并及时调整提示词。根据每次迭代的反馈结果,优化提示语和参数设置,以提高大语言模型知识抽取的准确度和处理效率。在知识图谱的构建过程中,提示工程与其他步骤(如数据预处理、知识融合、存储和验证等)须紧密配合。在提示工程中,应确保其与各个阶段无缝衔接,使每一轮抽取与优化都能顺利融入知识图谱构建过程,形成连续有效的知识链条。最后,需要对采用的方法与流程进行文档化保存,以保障提示工程的可持续性。通过这些步骤,可以确保大语言模型能够有效地抽取和理解文本中的复杂信息,通过反复进行提示语优化与参数调整,最终确保大语言模型能够高效、准确地处理传统文化文本中的复杂信息,为山水志知识图谱构建提供坚实的基础。在提示工程的设计部分,文章根据不同的实体、属性和关系提取设计了具体

通过这样的设计,能够确保模型在实际应用中精准地识别和抽取出所需的文化信息。

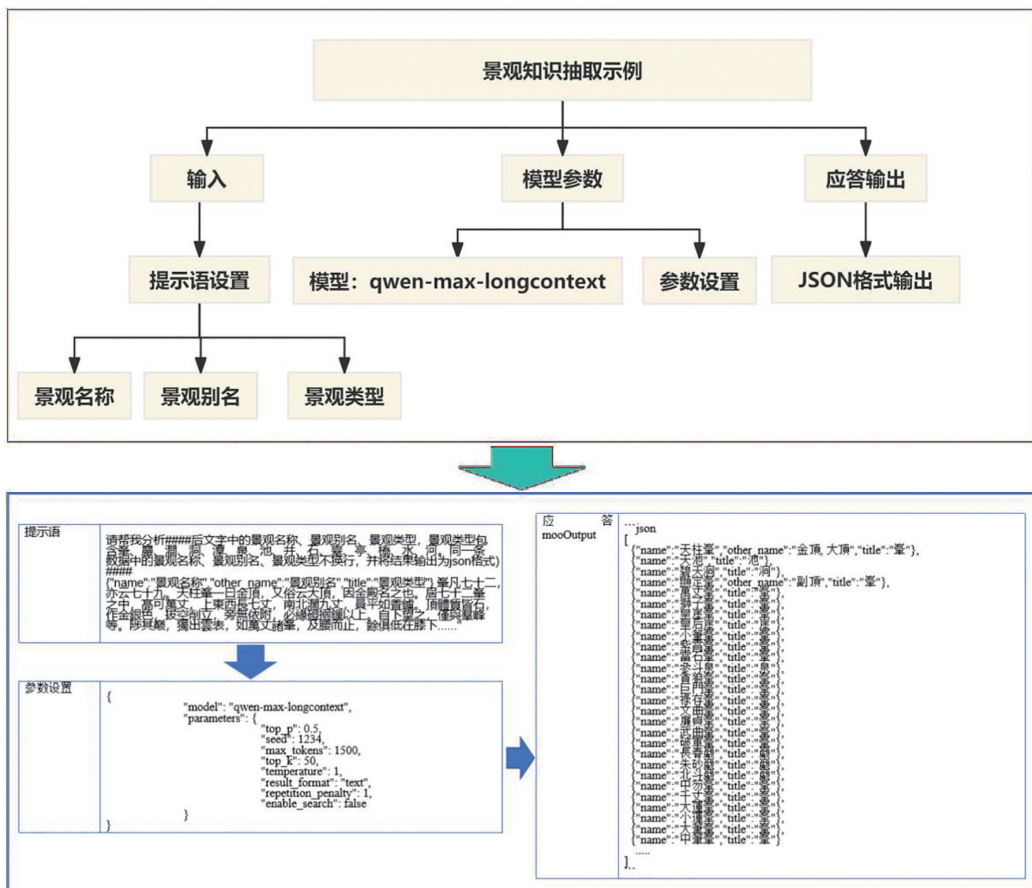


图2 基于提示工程的景观知识抽取示例

采用人工干预的方式,通过专家验证和修正异常数据,确保最终结果的准确性。完成后,对图谱进行验证,以确保其信息的准确性和可靠性,最后将验证后的知识存储在图数据库 Neo4j 中。在此基础上,知识图谱需要定期更新和维护,并通过应用和服务将其提供给用户。

以山水志知识图谱为例,该构建过程可以分解为多个子流程,逐步完成各类知识信息的抽取。具体如图3所示,景观实体与属性、景观人物关系以及人物实体与属性的提取都可以通过一系列算子模块处理后,分别将景观、人物和关系数据存入图数据库Neo4j中。文本在经过输入和读取后,景观实体属性数据通过提示语、通义千问大语言模型、值抽取等模块存入数据库,景观人物关系与人物属性的抽取则与此类似,区别在于对提示词的定義及属性、关系值的设定。通过此流程可一次性完成对任意输入文本中人物、景观和关系知识的构建,为研究者和用户提供一個全面、准确且可维护的知识图谱。

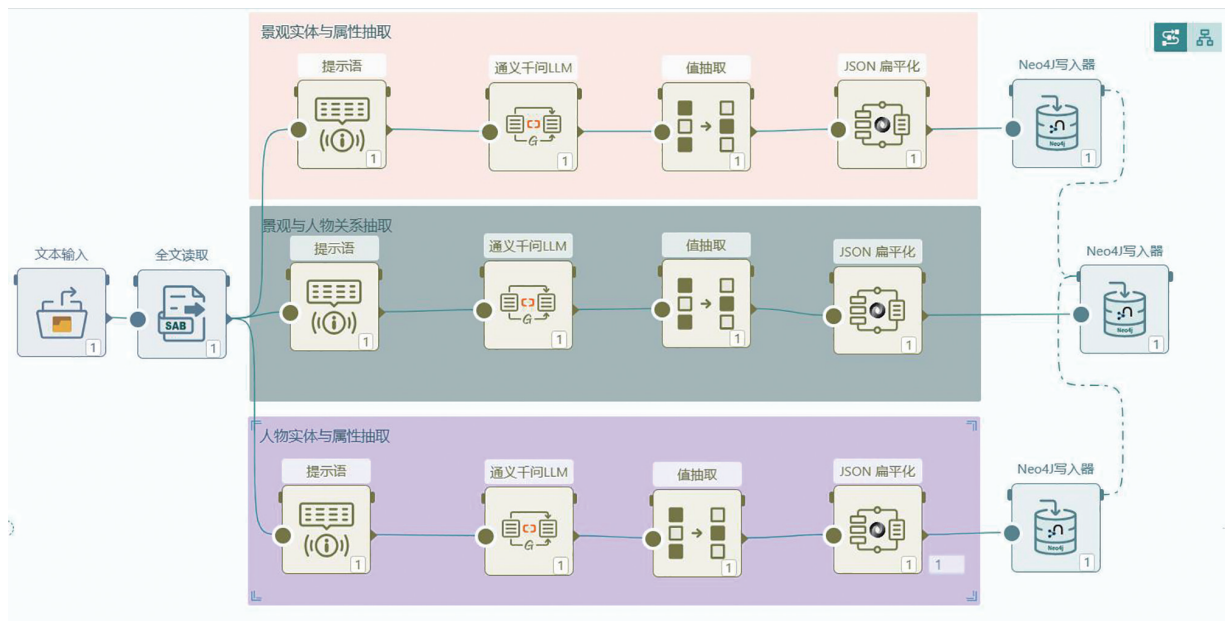


图3 面向人物、景观要素的知识图谱构建流程

通过评估大语言模型在实体识别、属性抽取、关系识别等任务中的表现,确保每个任务能够精确地处理山水志中的特定语言结构和文化背景。准确性检验方面,为了评估大语言模型在知识抽取中的准确性,文章采用了人工标注的金标准数据集,计算得到大语言模型在实体关系抽取中的准确率均在90%以上。结果表明,大语言模型能够有效地识别山水志中的核心实体和它们之间的关系,展示了模型在处理特定领域文本时的较高准确性。一致性检验方面,文章对相同的山水志段落进行了多次抽取实验。结果显示,大语言模型在实体、属性、关系等元素的识别方面具有较高的一致性,重复实验中关系抽取的标准差仅为5%,表现出稳定性。领域适用性检验方面,在跨文本验证过程中,模型成功地在《续修大岳太和山志》和现有数据集《沙湖志》^[5]之间实现了迁移学习,并且在两种文献的实体与关系抽取一致性达到了92%以上,表明了大语言模型能够较好地适应山水志等地方志领域的特定文本。最后,通过人工复核,团队对大语言模型的抽取结果进行了人工验证,确保其符合历史事实,并对模型的误差进行了修正。

3.4 山水志史料知识可视化与知识发现

山水志史料知识具有多维性特征,其知识单元通过语义关联的构建实现了多维度的知识聚合与重组,从而带来知识的增值与再生产。文章基于知识

图谱技术重构了武当山水文化记忆的语义关联,探讨了跨时空的文化记忆场景重构路径,实现了山水文化记忆的多层次呈现与语义化分析。

3.4.1 知识可视化:语义网络展示

知识图谱的核心功能之一是通过可视化展示复杂的语义网络,直观揭示文化记忆中不同知识单元的关联。《续修大岳太和山志》(以下简称《山志》)详细记录了武当山的地理位置、自然景观与人文历史,作为其知识可视化的基础数据来源,展现了武当山在山水文化记忆中的独特地位。

武当山位于荆、梁、豫三地交界,其地理位置和自然景观在《山志》中被高度重视。《山志》详细记录了太和山的七十二峰、三十六岩、二十四涧等景观体系,这些景观类型包括峰、岩、涧、洞、潭、泉、桥等,每一景观都根据其地形特征命名,例如黄龙洞因形似黄龙而得名,虎耳岩因形似虎耳而得名。知识图谱通过整合这些景观与相关描述,构建了武当山自然景观的语义网络。除了自然景观,《山志》还记载了武当山的宗教文化景点和历史人物活动。例如,马明生在炼丹池炼制“大日丹”,以及陈希夷的隐居活动。图谱进一步关联了与这些景观相关的重要历史事件,如尹喜隐居老君洞的传说,这些信息通过知识图谱的语义网络展示得以清晰呈现(见图4和图5)。通过知识图谱,研究者不仅能够观察景观的个体特性,还能直观理解它们在整个文化体系中的意义和作用。

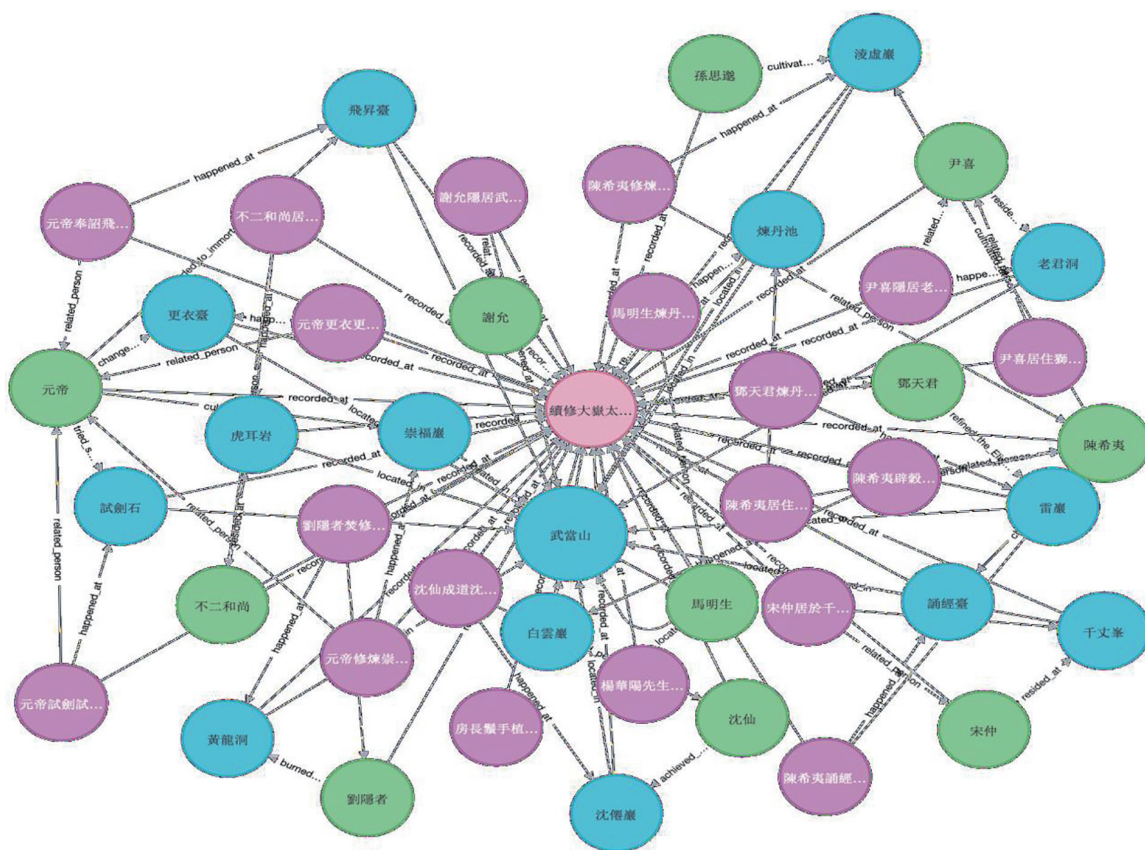


图4 武当山山水胜迹的知识可视化(部分)

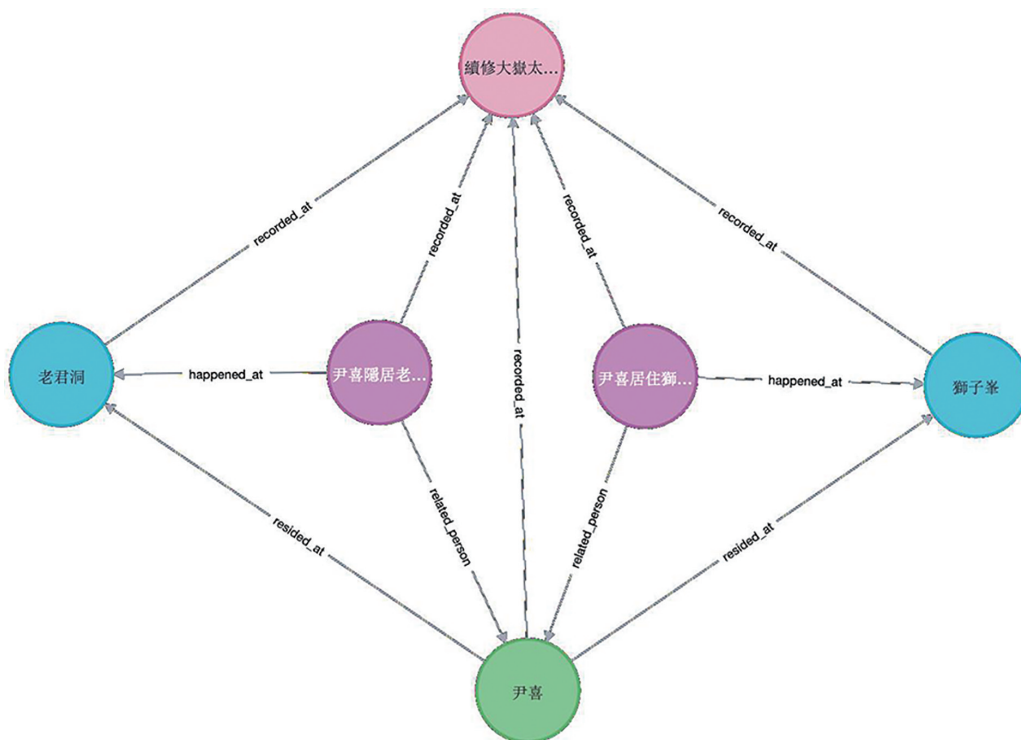


图5 武当山尹喜相关事件的知识可视化



3.4.2 语义分析:隐性知识发现

知识图谱不仅能用于可视化,还能够通过语义化分析揭示隐性知识,为研究者提供新的认知路径。以武当山为案例,文章基于知识图谱的语义分析功能,探讨了真武信仰的传播、历史事件的文化意义以及山水文化的记忆维度。

武当山作为道教圣地,其文化记忆的核心在于真武信仰的传承与发展。这一信仰自宋代起获得了朝廷的官方认可,并在明代达到鼎盛,通过系统化的宗教仪式和庙宇建设得到巩固。例如,在宋真宗天禧二年(1018年),真武被封为“镇天真武灵应祐圣真君”;明嘉靖三十二年(1553年),真武宫重修并由皇室派遣张彦主持祭祀;清康熙年间,真武被加封为“北极镇天真武祖师万法教主元天元圣仁

威上帝”。这些宗教事件被以时间轴的形式整合到知识图谱中,如图6所示,通过语义分析可以直观呈现真武信仰在历代王朝推动下的传播与发展脉络。此外,知识图谱的语义分析功能还揭示了历史人物在武当山文化记忆中的贡献。例如,通过关联尹喜的活动记录与景观节点(如老君洞),研究者可以进一步探讨其隐居生活对道教文化和武当山地位的影响。类似地,通过分析文化活动与景观节点的关联,可以发现宗教活动如何塑造和深化了武当山的文化身份。

从整体上看,知识图谱不仅为《山志》的多维知识单元提供了可视化工具,还通过语义分析帮助研究者挖掘出文本中隐含的知识关联,为武当山文化记忆的数字化保护提供了新方法。

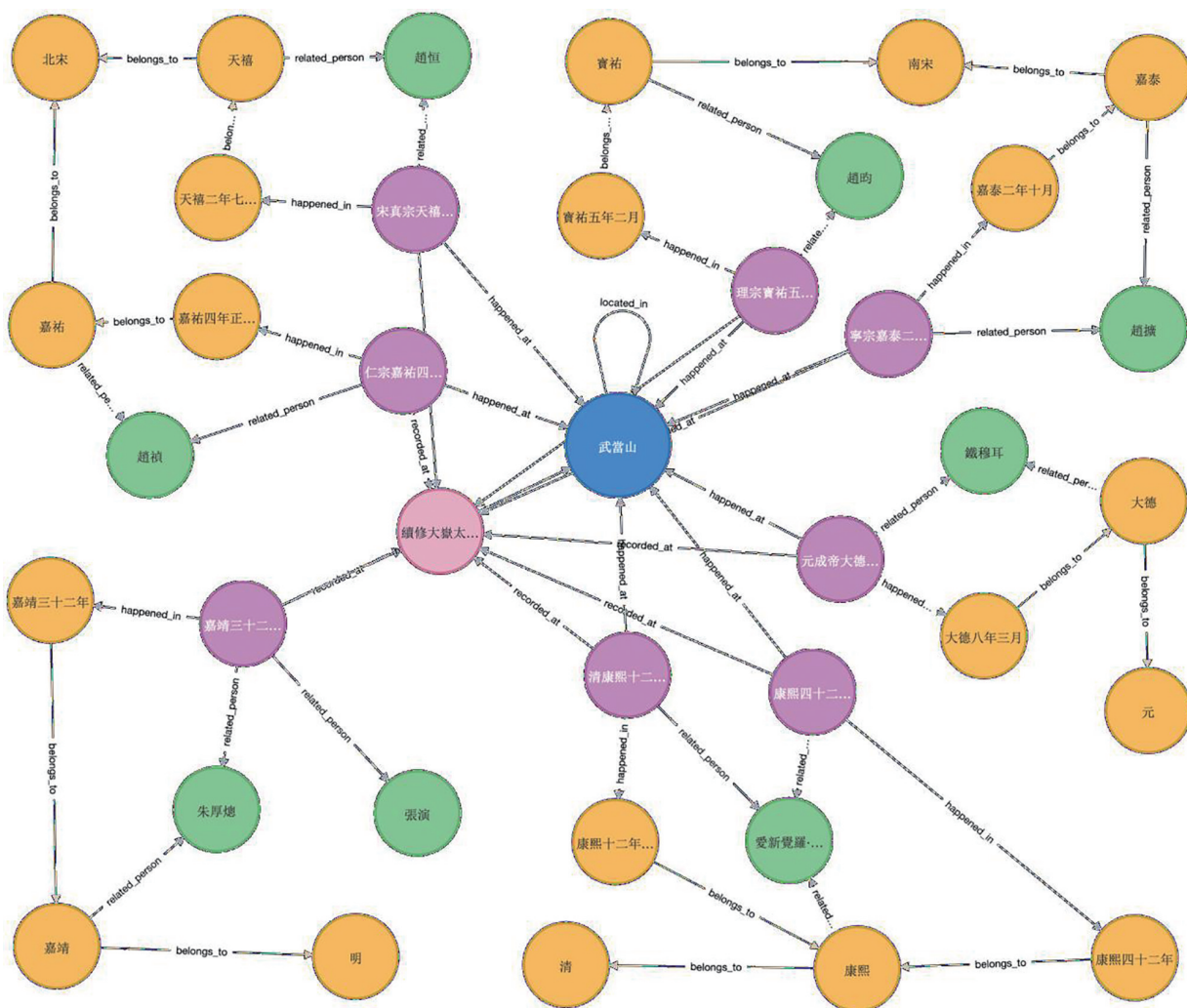


图6 武当山真武宗教活动知识图谱



4 结语

中国山水志史料是客观记录地方地理环境、社会生活的重要文献资源,同时也是挖掘地区文化传承脉络与历史演进规律的核心依据。文章以山水志为研究对象,选取武当山志作为实证案例,以“物质、象征、功能”三个维度为研究主线,融合大语言模型进行知识抽取的基础上构建了山水文化知识重构路径。通过知识图谱的构建与可视化展示,实现了从山水志文本数据中抽取、整合到构建语义化知识网络的完整流程,将传统史料中静态化、扁平化的知识转化为立体化、动态化的语义网络。最终,文章探索了山水志史料资源在知识发现、语义分析与场景化利用方面的潜力,为山水文化的活化利用和价值提升提供了新的应用路径。

文章的创新点主要体现在理论与实践结合、技术创新应用、知识可视化与发现三个方面。首先,通过记忆之场理论与山水文化“形象意”三维特征的结合,构建了物质、象征、功能三维度驱动的知识重构框架;其次,以大语言模型与低代码平台 Huggingfists 为基础构建垂直领域知识图谱,显著提升了知识抽取的效率,降低了技术门槛,使得非技术型研究人员能够更加便捷地进行知识图谱的构建;最后,基于 Neo4j 图数据库构建了动态知识图谱,不仅实现了知识的语义化关联,还通过可视化与语义分析工具支持隐性知识的挖掘与文化传播规律的研究。尽管文章在山水志知识图谱构建方面取得了一定成果,但仍存在一些值得进一步探索与优化的方向。首先,文章主要集中于知识图谱的基本构建,未来研究可以进一步细化实体属性,扩展关联关系,并探索基于 GraphRAG(图谱增强生成式模型)的方法,以更全面、深入地呈现山水文化的知识体系。其次,知识图谱构建后如何更好地应用于文化遗产保护、教育推广和公众传播等实际场景,也是需要重点解决的问题。

参考文献

- 1 闫慧. 文化数字化发展政策比较研究[J]. 情报理论与实践, 2022, 45(8): 9-14, 30.
- 2 中国政府网. 中共中央办公厅 国务院办公厅印发《关于推进实施国家文化数字化战略的意见》[EB/OL]. [2025-04-20]. https://www.gov.cn/xinwen/2022-05/22/content_5691759.htm.
- 3 黎安润泽, 牛力, 郑金月. 解构与活化: 基于双视角的地方志多粒度知识服务模型[J]. 图书馆杂志, 2024, 43(8): 111-122.
- 4 Bao P, Zhu S. System design for location name recognition in ancient local chronicles[J]. Library Hi Tech, 2014, 32(2):

- 276-284.
- 5 高劲松, 周树斌, 高颖, 等. 山水志史料资源语义知识关联与多维知识发现研究[J]. 情报资料工作, 2023, 4(5): 82-92.
- 6 Burdick A, Drucker J, Lunenfeld P, et al. Digital humanities[EB/OL]. [2025-04-20]. https://archive.org/details/DigitalHumanities_201701.
- 7 张强, 吴艳飞, 高颖, 等. 基于知识图谱的黄河流域非遗资源智能问答研究[J]. 文献与数据学报, 2023, 5(3): 100-115.
- 8 熊宾. 续修大岳太和山志[EB/OL]. [2025-04-20]. http://read.nlc.cn/allSearch/searchDetail?searchType=12&showType=1&indexName=data_403&fid=312001076766.
- 9 冷雪卓, 张涛, 崔文波. 国内数字人文领域文献主题识别及框架研究——基于 LDA 和 Word2Vec 的分析[J]. 图书馆学刊, 2024, 46(3): 104-110.
- 10 韩春磊, 徐卓韵. 知识图谱赋能下的历史人物数据平台构建实践——以宋庆龄文献数据中心平台建设为例[J]. 图书馆杂志, 2024, 43(6): 114-123.
- 11 张云中, 汪诗雨. 基于知识图谱的井冈山革命文化资源数字叙事模型的构建与实现[J]. 图书情报工作, 2024, 68(23): 106-119.
- 12 谢玮, 杨家瑶. 以“耕织图像”为核心的知识图谱设计及应用研究[J]. 大学图书馆学报, 2024, 42(1): 83-92.
- 13 沈雪莹, 欧石燕, 卢彤彤. 中国古代文人生平知识图谱构建与应用: 以李白和杜甫为例[J]. 数字图书馆论坛, 2023(8): 1-14.
- 14 张云中, 孙平. 历史文化名人游学足迹知识图谱的构建与可视化[J]. 图书馆杂志, 2021, 40(9): 81-87.
- 15 卢彤彤, 欧石燕, 李昕尉, 等. 古代战争知识图谱构建与应用——以宋元时期为例[J]. 图书馆论坛, 2024, 44(8): 90-101.
- 16 李永卉, 刘沁芃, 卢章平. 基于知识图谱的南朝陵墓石刻信息资源开发研究[J]. 图书馆杂志, 2023, 42(9): 86-93, 129-135.
- 17 熊欣, 王昊, 邓三鸿. 面向方志知识图谱的术语抽取模型迁移学习研究[J]. 情报理论与实践, 2021, 44(4): 176-184.
- 18 徐晨飞, 包平, 张惠敏, 等. 基于关联数据的方志物产史料语义化知识组织研究[J]. 大学图书馆学报, 2020, 38(6): 78-88.
- 19 徐晨飞, 包平. 基于方志物产知识库的数字人文研究框架设计与实证研究[J]. 中国科技史杂志, 2022, 43(2): 218-233.
- 20 徐晨飞, 唐佳林. 方志书目提要语义化知识组织与知识发现研究[J]. 数字图书馆论坛, 2024, 20(11): 30-42.
- 21 张兴旺, 李洁, 李思凡, 等. 基于大语言模型的古籍文献知识组织框架构建及其应用研究[J]. 图书馆, 2025(4): 7-12.
- 22 陈和. 基于大语言模型的参考文献自动识别与著录信息抽取[J]. 情报杂志, 2025, 44(7): 192-198.
- 23 陈华钧. 大模型时代的知识处理: 新机遇与新挑战[EB/OL]. [2024-09-12]. <https://www.secrss.com/articles/58455>.
- 24 孙翌, 刘音. 基于知识图谱和大语言模型的口述历史资源的问答应用研究[J]. 图书馆杂志, 2025, 44(1): 98-107, 119.
- 25 周正达, 王昊, 汪琳, 等. ChatKG 一种基于大语言模型和提示工程的非遗知识图谱构建框架——以中国非遗陶瓷制作工艺为例[J/OL]. 图书馆杂志, 1-30 [2025-09-02]. <https://link.cnki.net/urlid/31.1108.g2.20250224.1136.008>.
- 26 李成晴. 山水志宋元诗文献价值考述[J]. 南都学坛, 2015, 35(3): 45-48.
- 27 石光明. 明清时期山水志书的学术价值研究[J]. 农业考古, 2006(1): 223-225.
- 28 皮埃尔·诺拉. 记忆之场——法国国民意识的文化社会史[M]. 黄艳红, 等, 译. 南京: 南京大学出版社, 2015: 76.



- 29 李姗姗,张鑫雨. 记忆之场理论视域下红色档案资源戏剧化开发模式与路径[J]. 山西档案, 2023(1): 5-13.
- 30 Künzler S. Sites of memory in the Irish landscape? Approaching ogham stones through memory studies [J]. Memory Studies, 2020, 13(6): 1284-1304.
- 31 吕妍,肖竞. 传统村落“记忆之场”识别与多情景保护更新研究[J]. 城市与区域规划研究, 2023, 15(2): 39-52.
- 32 李姗姗,崔璐,崔旭. 记忆之场理论视域下档案史料的解构与重构[J]. 档案学研究, 2023(2): 103-110.
- 33 卢春红. 意象的山水化及其艺术表达[EB/OL]. [2025-09-12]. https://www.cssn.cn/skgz/bwyc/202507/t20250714_5886047.shtml.
- 34 王概. 大岳太和山纪略[EB/OL]. [2025-04-20]. http://read.nlc.cn/allSearch/searchDetail?searchType=12&showType=1&indexName=data_403&fid=312001076765.

作者贡献说明:

周树斌:提出选题,设计研究思路与框架,论文撰写及修改

程双玲:数据调研与分析,论文撰写及修改

作者单位:周树斌,内蒙古师范大学科学技术史研究院,内蒙古自治区科技文化遗产认知智能重点实验室,内蒙古呼和浩特,010022

程双玲,庆北大学图书情报系,韩国大邱广域市,41566

收稿日期:2025年5月16日

修回日期:2025年9月12日

(责任编辑:李晓东)

Research on Knowledge Reconstruction and Visualization Path of Ancient Chinese Landscape Chronicles Through the Integration of Large Language Models

—A Case Study of the *Continuation of the Great Yue Mount Taihe Chronicle*

ZHOU Shubin CHENG Shuangling

Abstract: As an important repository of Chinese landscape culture, landscape chronicles serve as a significant cultural memory carrier. Exploring the pathways for knowledge reconstruction and visualization with the aid of digital intelligence technologies can further enhance their cultural and mnemonic value. This study focuses on landscape chronicles materials and designs a knowledge graph construction plan based on three semantic dimensions: material, symbolic, and functional. It proposed a research pathway for knowledge reconstruction and visualization, covering four key modules: resource collection, knowledge representation, knowledge extraction, and knowledge visualization. Incorporating the theory of “Sites of Memory,” the study deepened the deconstruction and reconstruction of the cultural memory and natural landscapes embedded in the landscape chronicles. This study used the *Continuation of the Great Yue Mount Taihe Chronicle* as a case study to conduct an empirical analysis. The entire process of constructing a knowledge graph was implemented through the integration of a large language model and the low-code platform Huggingfists. First, knowledge extraction from landscape chronicle texts was performed through designed prompts. Entities, attributes, and relationships within the text dataset were identified based on standardized conventions for knowledge representation. Subsequently, the extracted discrete knowledge was integrated, and semantic associations between entities, attributes, and relationships were established through normalized descriptions of knowledge representation. Finally, the extracted and structured landscape cultural knowledge was stored in Neo4j as triples. The dynamic knowledge graph built on the Neo4j graph database visually illustrated the interconnections between Wudang Mountain’s core landscapes, associated historical events, and cultural symbols. This multi-perspective, multi-level visualization enhanced the intuitiveness of landscape knowledge and provided comprehensive semantic context for cultural research. Leveraging semantic analysis tools, it supported the exploration of implicit knowledge and the study of cultural dissemination patterns. In summary, this study achieves a complete workflow from extracting and integrating data from landscape chronicle texts to constructing a semantic knowledge network through knowledge graph construction and visualization. It transforms static, flat knowledge in traditional historical materials into a three-dimensional, dynamic semantic network. This approach offers valuable exploration pathways for the digital development of landscape chronicle materials and the preservation of landscape culture. It also provides a feasible solution for integrating large language models into the fields of history and the humanities.

Keywords: Landscape Chronicle; Knowledge Graph; Large Language Model; Knowledge Reconstruction; Digital Humanities