

·业务研究·

基于大语言模型的高校图书馆传统文化阅读推广的知识图谱构建及优化路径

王丽丽

(吉林大学 图书馆,吉林 长春 130012)

摘要:【目的/意义】在新时代背景下,探索基于大语言模型的高校图书馆传统文化阅读推广的知识图谱构建流程,通过可视化图谱优化阅读推广路径,推动传统文化资源的深度挖掘与系统化组织,创新高校图书馆的阅读推广模式。【方法/过程】本研究提出一种基于大语言模型的传统文化阅读推广内容的知识图谱构建流程,采集数据后先进行数据清洗处理,运用BERTopic主题模型提取实体类别与关系标签,再使用deepseek-r1大语言模型进行实体抽取与关系识别,最后将实体关系三元组导入Neo4j图数据库,以实现知识图谱的可视化展示。此外并利用百度搜索以及微信搜一搜平台中的具体内容构建高校图书馆传统文化阅读推广知识图谱,探讨其在传统文化阅读推广中的创新路径模式。【结果/结论】研究结果表明:基于deepseek-r1模型的知识图谱构建方法能够清晰呈现传统文化阅读推广中的组织机构、活动形式及推广策略等,形成结构化的知识表达体系;进一步分析传统文化阅读推广知识内容,可辅助传统文化阅读案例库建设,助力构建智能化问答系统,协助挖掘传统文化阅读推广路径,以发挥高校图书馆文化传承的阵地作用。【创新/局限】提出了一种基于大语言模型的传统文化阅读推广内容的知识图谱构建方法,并挖掘知识元素用于高校图书馆传统文化推广,但由于计算机资源的限制,未来的研究中可以借助第三方平台的大模型接口。

关键词: 大语言模型;BERTopic;知识图谱;传统文化;阅读推广

中图分类号: G252.2 **DOI:** 10.13833/j.issn.1007-7634.2024.12.015

0 引言

中华优秀传统文化是中华民族在漫长历史进程中积累的宝贵精神财富,也是新时代繁荣和发展社会主义先进文化的重要支撑,具有交融性、德育性、稳定性等特征^[1]。在全球文化传播和数字化信息快速推进的背景下,传统文化的传承与推广显得愈加重要,高校图书馆承担着保护、传承文化遗产和发扬中华优秀传统文化的重要职能^[2-3],然而,由于传统文化的资源分散、信息组织和推广方式的传统局限性,使得高校图书馆在履行这一职能时面临诸多挑战。

为应对这些挑战,优化高校图书馆的传统文化阅读推广路径成为关键策略,以期为高校图书馆传统文化阅读推广创新提供参考,促进传统文化知识的传播。鉴于此,本文将目光投向了新兴的大语言模型技术,特别是以大语言模型的本地部署调用和提示词工程为基础的传统阅读推广知识图谱的构建。传统文化阅读推广知识图谱是一种图形化的知识组织形式,能够清晰展现传统文化阅读推广活动之间的复杂关系^[4]。通过实体识别、关系抽取等技术手段,图书馆中的文化资源,推广活动的目标、宣传手段、活动内容等关键信息被转化为节点和边,从而构建出一个多维度、动态更新的知识网络^[5]。而本地部署的大语言模型作为一种高效的

自然语言处理工具,凭借其强大的理解和生成能力^[6],能够自动化地分析传统文化阅读推广文本并提取关键信息,为知识图谱的构建提供了坚实的技术保障。此外,提示词工程通过精心构造输入提示,有效地引导模型执行任务,提升结果的准确性,在传统文化阅读推广知识图谱的构建过程中,合理运用提示词设计能够进一步优化信息提取的效果和精度。

本研究基于大语言模型,提出了一种优化的传统文化阅读推广知识图谱构建流程,通过数据清洗和BERTopic主题模型的实体类别与关系标签提取,结合deepseek-r1大语言模型进行实体抽取与关系识别,最终将实体关系三元组导入Neo4j图数据库,实现知识图谱的可视化展示。并以百度、搜狗、微信和秘塔AI平台内容为研究对象,探索构建传统文化阅读推广知识图谱,期望通过本研究,为高校图书馆的传统文化阅读推广提供一条创新的技术路径,推动文化资源的深度挖掘与高效利用,助力文化传承与学术研究。

1 相关文献综述

1.1 传统文化阅读推广的相关研究

传统文化阅读推广是推动全民阅读深化与传统文化传承的重要实践路径,融入高校图书馆服务体系后,能够有效

收稿日期:2024-06-07

作者简介:王丽丽(1982-),女,吉林长春人,硕士,馆员,主要从事知识服务,信息服务研究。

- 130 -

促进文化资源整合与推广策略创新,增强受众群体的文化认同与参与黏性,助力构建可持续的传统文化传播生态体系^[7]。当前相关研究聚焦于推广策略创新、技术融合、群体适配与效果评估维度。在策略与实践层面,学者强调内容多元化和模式创新;刘梦佳^[8]提出高校图书馆需通过数字化升级与长效机制破解内容单一困境;穆允军等^[9]基于古典目录学理论构建古籍资源整合框架,推动系统性推广。数字化与新媒体融合研究凸显技术赋能价值,崔钰琪和杨红英^[10]主张构建民族传统体育文化“线上+线下”数字空间,强化资源与活动联动;黄运红等^[11]借助符号学理论设计“激发—推动型”双主体互动模式,提升新媒体场景的象征性与具身体验。

特殊群体推广与长效评估机制成为研究新焦点。针对青少年及流动儿童,汪全莉等^[12]构建阅读活动生命力评估体系,提出全周期优化策略;陈筱琳^[13]基于苏州流动儿童“中国艺术课”项目,证实传统文化阅读对心理适应性 with 学习动力的正向干预。评估体系研究方面,孙媛媛等^[14]依托“二十四节气”立体阅读案例,提出效果评价与模式优化的闭环逻辑;伍力^[15]从资源整合、传播协同与多方合作三维度,构建公共图书馆传统文化推广体系。

1.2 大语言模型与知识图谱构建及相关研究

近年来学者们运用提示工程、微调技术、多模态数据融合等大语言模型方法,对不同领域内知识图谱构建展开研究,相关研究聚焦于构建优化、推理补全、技术协同与垂直应用四大方向。在知识图谱构建与扩展领域,学者通过 LLM 技术显著提升知识提取效率与精度;黄俏娟等^[16]基于 GPT 系列模型设计事件知识提示工程,将中文常识知识图谱扩展至 243 万条,时序关系三元组准确率达 98.2%;冯钧等^[17]针对水工程调度领域,利用 LLM 从非结构化文本提取知识构建领域 KG (F1=85.5%),强化水利规则结构化能力。知识推理与补全研究则凸显 LLM 的语义理解优势,陈娟等^[18]提出“故事启发”时序推理方法,通过 LLM 生成叙事逻辑增强稀疏事件关联预测;杨燕等^[19]构建 KG-LLM 融合模型,在虚假健康信息识别中实现 98.35% 准确率,有效抑制 LLM 幻觉问题。

模型优化与垂直应用研究进一步拓展技术边界:例如司悦航等^[20]提出时序 KG 蒸馏框架,融合领域时序特征实现轻量化推理模型部署;张琪等^[21]通过次模优化生成三元组语义嵌入,结合贪心算法提升实体摘要质量 (NDCG 指标优于基线)。在垂直领域应用中,张鹤译等^[22]构建中医药问答系统,通过 LLM 与 KG 协同生成专业回答;张学飞等^[23]设计教育领域个性化推荐系统,利用 KG 表示学习路径与 LLM 动态生成资源,解决冷启动与可解释性问题。

1.3 相关研究述评

综上所述,现有研究存在一定的局限性,特别是在传统文化资源的整合与共享方面,推广内容的多样性和跨领域的适用性较弱。因此,大语言模型 (LLM) 与知识图谱 (KG) 为

传统文化阅读推广提供了新的突破口,LLM 具有强大的文本理解与生成能力,能够从大规模非结构化数据中提取知识,而知识图谱则通过结构化的信息展示,帮助实现资源之间的关联与优化。构建基于大语言模型的传统阅读推广知识图谱,能够有效整合多元化的资源,提升内容推送的精准度,并通过知识图谱的可视化展示,优化推广路径,解决现有推广模式中的不足^[24]。

鉴于此,本文将基于大语言模型构建传统文化阅读推广的知识图谱,并利用其强大的推理与语义理解能力,通过 Neo4j 图数据库实现知识图谱的可视化,推动高校图书馆传统文化资源的深度挖掘与系统化组织,帮助制定更加精准的阅读推广路径。同时,借助知识图谱,研究将探索如何将其应用于智能化问答系统和个性化推荐服务中,从而推动高校图书馆服务的智能化转型与数字化升级。

2 传统文化阅读推广知识图谱的构建流程

本文提出一种传统文化阅读推广的知识图谱构建方法,先在百度搜索、微信公众号搜一搜等平台搜集并编写数据采集程序得到此主题下的数据集;然后,运用 deepseek-r1 大模型,构建高校图书馆传统文化阅读推广知识图谱;再基于知识图谱探讨其在高校阅读推广的创新应用路径,构建流程如图 1 所示。

2.1 数据采集与预处理

通过使用 Python 的 Scrapy 与 Selenium 模块, RPA 自动化程序,从百度搜索、微信搜一搜等平台获取各个图书馆关于传统文化阅读推广的相关内容,汇总得到文本数据集,内容覆盖关于高校图书馆的阅读推广模式、推广活动、数字资源和平台工具多个维度,以确保数据本身的多样性和丰富性。

针对采集下来的文本数据集进行预处理,经过数据清洗、降噪处理、规范化合并、文本预处理等步骤,并结合 deepseek-r1 的 API,清理规范文本数据,其中,数据清洗主要用于去除重复记录和缺失值,确保数据的完整性;降噪处理通过过滤网页广告、无效内容及短文本,剔除无关信息以减少干扰;规范化合并则统一不同的数据格式,规范数据结构,以便于 deepseek-r1 分析;基本的文本预处理的操作,例如切词分词、去停用词、保留同义词等过程,用于处理输入 BERT-Topi 主题模型进行分析,以进一步提取实体关系的类别标签。

2.2 传统文化阅读推广知识图谱实体关系抽取方法

在知识图谱构建阶段,首先利用 BERTopic 主题模型对文本数据集进行主题分析,在各主题中概括出知识图谱的实体关系标签;然后通过编写调用 deepseek-r1 模型 API 的程序,逐个提取每条文本的实体关系,汇总所有文本的实体关系数据集,即汇总实体关系三元组;最后,将提取的实体及关系三元组导入 Neo4j 数据库进行存储,通过 Cypher 语句,

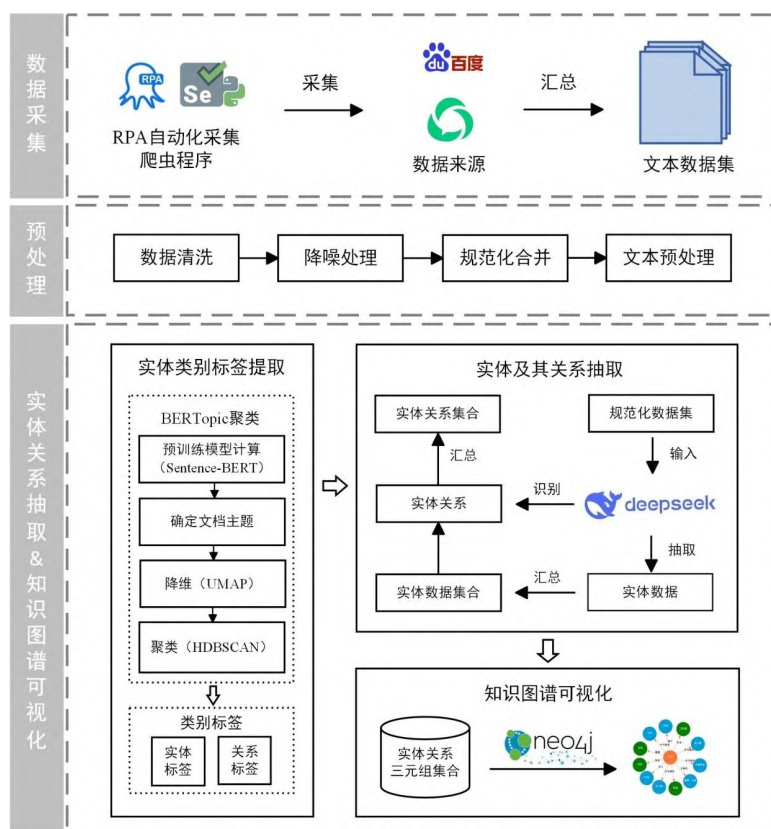


图1 传统文化阅读推广知识图谱构建流程

Figure 1 Process of constructing knowledge graph for traditional culture reading promotion

使高校图书馆传统文化的阅读推广的模式路径等内容结构清晰呈现^[25]。

2.2.1 BERTopic提取实体类别标签

BERTopic是一种主题模型^[26],结合Transformer^[27]和c-TF-IDF技术,通过构建密集的主题集群,不仅能够识别每个主题的关键特征,还能提高模型的可解释性,其基本步骤:(1)文档嵌入:BERTopic使用预训练语言模型,例如Sentence-BERT^[28],将文档转换为向量表示,使得计算机能够理解和处理;(2)降维:由于文档向量的维度很高,直接对其进行聚类会增加计算成本,故BERTopic采用UMAP算法^[29]对文档向量进行降维,提升计算效率。(3)聚类:在降维后使用聚类算法HDBSCAN^[30]对文档向量进行聚类,HDBSCAN能够自动确定聚类数量,并且处理噪声数据。(4)计算各主题中的每个词语权重值:利用c-TF-IDF算法计算每个类别中词语的权重,c-TF-IDF不仅能够衡量单词在特定主题中的重要性,还能剔除在整个语料库中高频但对某一主题贡献有限的词汇。(5)主题选择与汇总:在计算出每个簇的主题词后,利用最大边界相关MMR算法^[31]对主题进行筛选与优化,删除不相关主题,此外,可以通过调整相似度阈值合并相近主题,并设置阈值控制主题数量,从而进一步优化模型的可解释性。

利用BERTopic主题聚类方法,对所有文本数据进行主题分析,以全面挖掘数据的语义特征和主题结构,提取出核

心实体类别标签及其关系类型。BERTopic通过将文本转化为语义嵌入向量,在向量空间中表示文本的语义特征,并基于无监督聚类方法对具有相似语义的文本进行聚类归并,形成多个主题,结合每个主题的内部语义特征,提取具有代表性的实体类别标签,并概括这些类别之间的关系类型,这些类别将为后续的实体关系抽取提供明确的类别标签。

2.2.2 deepseek-r1模型

deepseek-r1是由深度求索研发的通用推理型大模型,核心创新在于通过多阶段强化学习(Reinforcement Learning, RL)框架激发语言模型的自主推理能力,区别于传统基于监督微调(Supervised Fine-Tuning, SFT)的范式^[32],deepseek-r1首次验证了无需人工标注数据、仅通过大规模RL训练即可使基座模型涌现出复杂推理行为,例如自我验证、反思式推理和长链思维过程,这些技术提升了模型在零样本任务上的性能,即使在未见过的数据上也能实现较高的精度和召回率,使得deepseek-r1在命名实体识别(NER)、关系抽取(RE)任务中具有独特的应用场景,成为构建知识图谱任务中的理想选择。

deepseek-r1模型包含原始的671B以及多个蒸馏版本,包括Qwen2.5和LLaMA系列,参数量在1.5B至70B,蒸馏模型虽然参数较少,但在性能上经过精心优化,能够在多种任务中提供高效的解决方案,适用于需要快速推理和计算资源限制的应用场景。因此通过部署蒸馏版本的大模型,调用

API的方法使其在构建知识图谱任务中的泛化能力有效支持其在零样本任务中的稳定表现。

2.2.3 实体及实体关系的抽取

利用 deepseek-r1 大模型实现实体及实体关系抽取,需要使用 ollama 从官网拉取 deepseek-r1 模型,通过代码调用 API 与 deepseek-r1 交互,其效果与 Prompt 的组织是紧密相关的,参考 ollama 源码的 github 网站,分别对 deepseek-r1 模型的 API 系统角色及提示词进行基础设置。

```
from ollama import Client

def get_completion(prompt, model_name= "deepseek-r1:32b"):

    client = Client(
        host='http://localhost:11434',
        headers={'x-some-header': 'some-value'})

    response = client.chat(model=model_name, messages=[{
        'role': 'user', # 用户角色
        'content': prompt, # Prompt 提示词
    }])

    return response['message']['content']
```

使用 Python 语言通过 ollama 模块与 deepSeek-r1 模型进行交互。首先,导入 ollama 库以便能够调用 API,Client 类变量存储了 API 的必要配置信息,这是使用 deepseek-R1 服务所必需的。在 get_completion 函数中,使用 client.chat 方法发送聊天信息给 deepSeek-r1 模型,该方法需要指定模型名称以及对话消息,对话消息设置了用户角色,构建一个模拟对话场景,角色的设置有助于提供上下文信息、模拟用户交互,并指导 DeepSeek-R1 模型进行更准确的响应^[5]。将 'user' 和 Prompt 内容整合为一个完整的 messages,使模型能够准确执行 NER 及 RE 的任务,并在传统文化阅读推广领域分析的相关数据并提取实体及关系。

3 传统文化阅读推广知识图谱构建的实证研究

3.1 实验环境简介

本实验的硬件环境:CPU,2*Intel(R) Xeon(R) Gold 6138 CPU @ 2.00GHz;内存,96GB;GPU,1*NVIDIA GeForce RTX 4090,显存 24GB。软件环境:操作系统 Ubuntu 24.04,Python 3.9.5、Pytorch 2.3.1、CUDA 12.2、Anaconda 3 和 PyCharm 2021。在此环境中,显卡资源足够部署 deepSeek-r1 的 32B 蒸馏模型。

3.2 数据获取与预处理

本研究通过 RPA 自动化与 Python 爬虫模块 Selenium 相结合的方式,数据来源为微信搜一搜和百度搜索;在百度搜索中,编写 Python 程序,使用 Selenium 库自动抓取与“传统文化阅读推广”相关网页内容,共计获取 723 张网页。在微信

搜一搜中,首先,借助秘塔 AI 联网查找出做传统文化阅读推广的高校图书馆典例,选择汇总 10 所高校图书馆,例如北京大学图书馆、四川大学图书馆、郑州大学图书馆等;接着,利用八爪鱼 RPA 逐个打开各个图书馆公众号,采集其中的所有文章;最后对每个图书馆公众号中的文章进行筛选,提取其中与阅读推广相关的文章,共计 466 篇,具体各所高校的文章数量如表 1 所示。

表 1 各高校图书馆公众号提取的文章篇数

Table 1 Number of articles extracted from the public libraries of each university

高校图书馆	文章篇数
北京大学图书馆	72
北京师范大学图书馆	65
四川大学图书馆	76
苏州大学图书馆	32
郑州大学图书馆	27
青岛大学图书馆	51
上海交通大学图书馆	43
浙江大学图书馆	56
南京大学图书馆	44

数据采集完成后,需要对原始数据集进行数据预处理,以提升数据质量和规范性^[33]。在数据清洗中,剔除网页数据中的重复记录和缺失值,确保数据的完整性和准确性;在降噪处理中,移除网页数据以及公众号文章中的广告、无效内容及短文本等噪声数据,优化内容的相关性和实用性;在规范化合并的过程中,利用 API 逐条处理,将网页数据与公众号文章中的文本内容规范化,Prompt 的结构如图 2 所示。经过系统化的数据预处理,最终获取到的传统文化阅读推广数据集包括 528 张网页与 423 篇公众号文章,数据样例如表 2 所示。

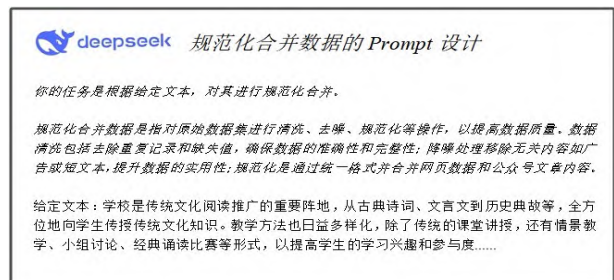


图 2 规范化合并数据的 Prompt 设计

Table 2 Prompt design for normalized merged data

3.3 实体关系抽取

3.3.1 实体关系标签提取

针对 528 张网页文本与 423 篇公众号文章,运用 BER-Topic 模型进行文本聚类,以期从不同主题中以获取实体及关系的类型标签。其中,在未确定主题数量 nr_topics 时,先调整参数使 nr_topics 为“auto”,即模型根据聚类结果自动确定主题数量,并设置邻近数据点 n_neighbors 为 15,数据点的

表2 传统文化阅读推广数据集样例
Table 2 Sample data set on traditional culture reading promotion

数据来源	具体内容
百度搜索	{ "title": "图书馆开展“中华优秀传统文化百部经典”阅读推广活动-图书馆”, "text": "为传承和弘扬中华优秀传统文化,认真落实国家图书馆、中国图书馆学会“2024年《中华优秀传统文化百部经典》阅读推广活动”安排部署,学校图书馆开展“览百部经典 践文化传承”——《中华优秀传统文化百部经典》阅读推广活动。通过主题展览、旅图悦影等形式,吸引广大读者深度阅读经典,领略传统文化魅力。\\n《中华优秀传统文化百部经典》系列图书所选典籍贯通古今,是中华优秀传统文化的思想精髓……” }
百度搜索	{ "title": "体验传统文化魅力,优质阅读为美好童年注入活力”, "text": "重温国学经典、制作非遗漂漆团扇、杯垫彩绘“二十四节气”、体验汉服汉礼、玩投壶、滚铁圈等游戏……日前,别开生面的“国学传承 童趣六一”——静安区少年儿童图书馆市集主题活动现场,百余名少年踊跃参与,感受传统文化的魅力。\\n每项沉浸式互动体验活动注入了浓厚的国学文化底蕴,让孩子们在寓教于乐的氛圍中“品尝”国学大餐,以童趣的方式触摸博大精深的中华优秀传统文化……” }
微信搜一搜	{ "title": "文脉相承,经典同行 北京大学2024年阅读文化节”, "text": "习近平总书记在文化传承发展座谈会上的重要讲话中指出,“只有全面深入了解中华文明的历史,才能更有效地推动中华优秀传统文化创造性转化、创新性发展,更有力地推进中国特色社会主义文化建设,建设中华民族现代文明。”为深入学习贯彻习近平文化思想,强化图书馆文化育人职能,2024年北京大学图书馆举办“文脉相承,经典同行:共读中华文明史”为主题的阅读文化节,通过讲座、书展、文化展、文化艺术体验、阅读报告发布等贯穿全年的系列活动,让不同学科的同学了解博大精深的中华文明,品读中国文学经典,坚定文化自信,增强文化自觉。……” }
微信搜一搜	{ "title": "期待你的声音 四川大学第六届“经典守护者” 中华经典美文诵读大赛火热征稿中~”, "text": "2023年是全面贯彻落实党的二十大的精神的开局之年,也是学校落实第九次党代会部署的关键之年。党的二十大描绘了全面建设社会主义现代化国家、以中国式现代化全面推进中华民族伟大复兴的宏伟蓝图。为深入贯彻党的二十大精神,落实立德树人根本任务,坚定文化自信,正确理解和大力推进中国式现代化,创新开展马克思主义经典和中华优秀传统文化经典诵读活动,着力建设书香川大,学校决定举办以“以青春助力中国式现代化”为主题的第六届四川大学“经典守护者”中华经典美文诵读大赛,建设四川大学校园诵读品牌,弘扬中华优秀传统文化……” }

排列密度 min_dist 为 0.1,共得到 26 个主题。调整目标维度数 n_components=2,绘制出降维后的主题距离图直观表征主题划分效果,如图 3 所示。

根据图 3,聚类结果存在部分主题重叠的情况,说明聚类效果还没有达到最优^[34],在第三象限与第四象限的距离较近两个类簇合并为一个,其余重叠类簇也均合并,重新确定 nr_topics=5,再 min_dist 为 0.2,重新训练 BERTopic 主题模型后的二维主题距离如图 4 所示。使用主题条形图直观展示每个主题维度下的一组关键词及其 c-TF-IDF 权重值,如图 5 所示。

依据主题提取的结果及可视化呈现,根据各个主题维度下的主题词可概括出五个维度的主题名称,分别为单位机构、活动形式、文化内容、推广策略以及受众群体,各主题名称分别对应于五种实体类型。在每类主题下分别根据 c-TF-IDF 权重值进行主题词筛选,梳理出最常出现的八种关系类型,具体实体类型与关系类型的逻辑连接表格如表 3 所示。

表3 实体类型与关系类型的逻辑连接表格

Table 3 Table of logical connections between entity categories and relationship types

主体类别	关系类型	客体类别
单位机构	举办	活动形式
单位机构	策划	推广策略

活动形式	应用	推广策略
活动形式	展示	文化内容
活动形式	吸引	受众群体
文化内容	支撑	推广策略
推广策略	依托	文化内容
受众群体	参与	活动形式

3.3.2 实体关系抽取 Prompt 设计

现有研究表明将任务解决步骤作为提示信息,以触发模型的思维链(Chain-of-Thought, CoT)是一种有效地提升大语言模型输出效果的提示优化方法^{[35][36]}。设计基于 CoT 提示优化的实体抽取的模板为: Prompt={Role, Demand, Schema, Form, Example, Text},其中,Role 为系统角色信息,提供明确的任务指引;Demand 为详细的任务说明信息;Schema 为五个实体类型的定义,通过详细说明各实体类型以帮助 deepseek-r1 模型识别不同类型的键信息;Form 为输出数据的格式要求;Example 为实体抽取的示例,使模型能够参考相似情境下的抽取方式,以提高抽取的准确性和一致性;Text 为需要抽取的网页文本或微信公众号文章。实体抽取的 Prompt 设计如图 6 所示。

同理,基于 CoT 提示优化的关系抽取的模板为: Prompt={Role, Demand, Schema, Form, Example, Text},其中,Role 为系统角色信息,提供明确的任务指引;Demand 为详细的任务说

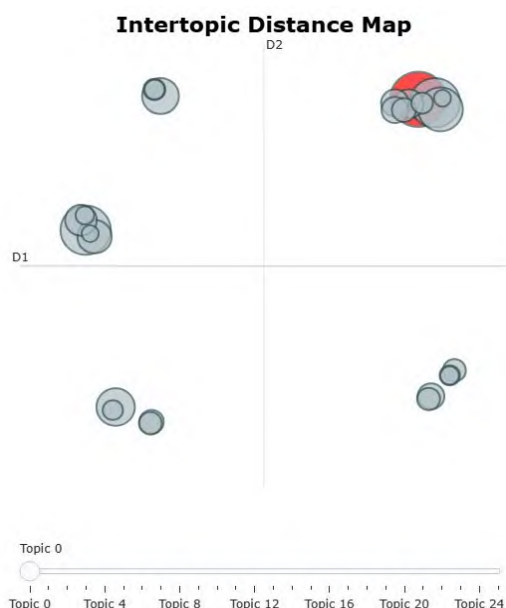


图3 nr_topics="auto"时聚类结果主题距离

Figure 3 Topic distance of clustering results with nr_topics="auto"

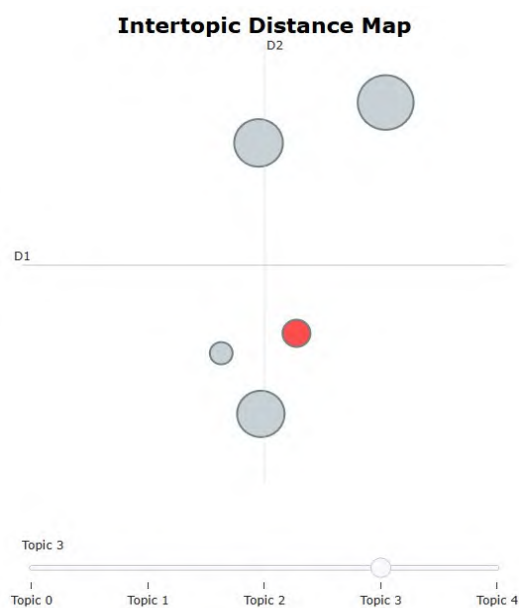


图4 nr_topics=5时聚类结果主题距离

Figure 4 Topic distance of clustering results with nr_topics=5



图5 BERTopic 聚类结果主题条形图

Figure 5 Thematic bar chart of bertopic clustering results

明信息;Schema为五个关系类型的定义,通过详细说明各关系类型以帮助deepseek-r1模型识别不同类型的键信息;Form为输出数据的格式要求;Example为关系抽取的示例,使模型能够参考相似情境下的抽取方式,以提高抽取的准确性和一致性;Text为需要抽取的网页文本或微信公众号文章。关系抽取的Prompt设计如图7所示。

3.3.3 实体关系抽取结果

通过利用实体抽取以及关系抽取的Prompt模板,将网页文本以及公众号文章逐条逐段输入到本地的deepseek-r1:32b模型进行实体抽取及关系抽取,并汇总最终抽取结果,共获得1178个有效实体和1562组有效实体关系对,部分实体的抽取结果及实体之间关系的抽取结果如下所示。

{"单位机构":["高校图书馆"],"活动形式":["阅读推广","专题问卷调查"],"文化内容":["中华典籍"],"推广策略":

["个性化内容","数字化技术","信息技术手段","交流互动阅读氛围","专业馆员培养"],"受众群体":["大学生"]}

{"单位机构":[],"活动形式":["座谈","经典诵读"],"文化内容":["中华优秀传统文化","经典书目"],"推广策略":["兴趣导向","生活应用"],"受众群体":["孩子","学生"]}

{"单位机构":["北京大学图书馆"],"活动形式":["讲座","书展","文化展","文化艺术体验","知识竞赛","图书交换活动","阅读马拉松","学生读书沙龙","线装书装订体验","腰封DIY体验","古诗吟诵","开幕式系列活动"],"文化内容":["史记","李太白集","文心雕龙","山海经","东坡集","花间词","穆桂英挂帅","茶馆","觉醒年代","西厢记","中华文明史","中国文学经典"],"推广策略":["阅读文化节","书香燕园","主题讲座","文化体验活动","知识竞赛","图书漂流活动","专题书展"],"受众群体":[]}



图6 实体抽取的 Prompt 模版设计
Figure 6 Prompt template design for entity extraction

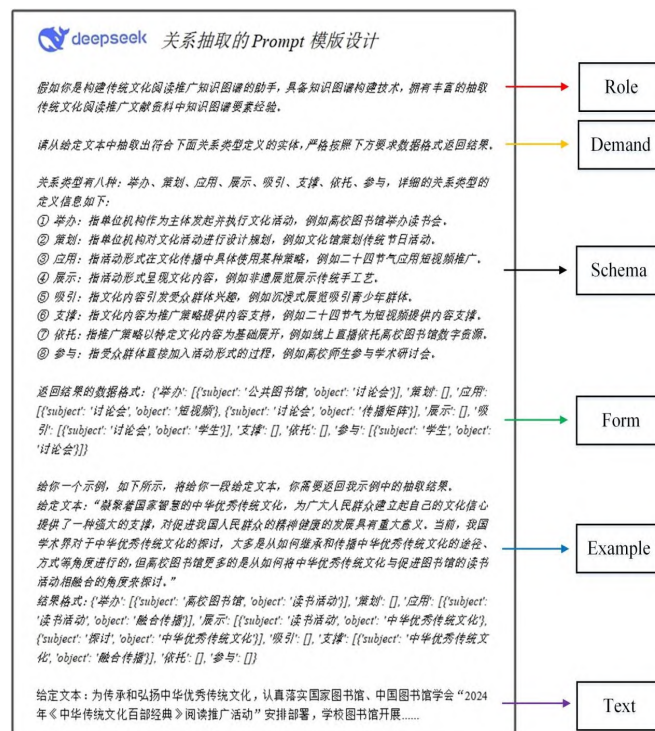


图7 关系抽取的 Prompt 模版设计
Figure 7 Prompt template design for relationship extraction

["举办": [{"subject": "高校图书馆", "object": "阅读推广"}], "策划": [], "应用": [{"subject": "阅读推广", "object": "数字化技术"}], {"subject": "阅读推广", "object": "信息技术手段"}], {"subject": "阅读推广", "object": "专业馆员培养"}], "展示

": [{"subject": "阅读推广", "object": "中华典籍"}], "依托": [{"subject": "个性化内容", "object": "中华典籍"}], "吸引": [{"subject": "阅读推广", "object": "大学生"}], "参与": [{"subject": "大学生", "object": "阅读推广"}], "支撑": [{"subject": "

中华典籍", "object": "个性化内容", {"subject": "中华典籍",
"object": "数字化技术"]]

{"举办": [], "策划": [], "应用": [{"subject": "经典诵读",
"object": "兴趣导向"}, {"subject": "经典诵读", "object": "生活应用"}], "展示": [{"subject": "经典诵读", "object": "经典书
目"}], "依托": [{"subject": "兴趣导向", "object": "中华传统文
化"}, {"subject": "生活应用", "object": "中华传统文化"}], "吸
引": [{"subject": "经典诵读", "object": "孩子"}, {"subject": "
经典诵读", "object": "学生"}], "支撑": [], "参与": [{"subject":
"孩子", "object": "经典诵读"}, {"subject": "学生", "object": "
经典诵读"}]}

{"举办": [{"subject": "北京大学图书馆", "object": "讲座
"}, {"subject": "北京大学图书馆", "object": "书展"}, {"sub
ject": "北京大学图书馆", "object": "文化展"}, {"subject": "北
京大学图书馆", "object": "文化艺术体验"}, {"subject": "北京
大学图书馆", "object": "知识竞赛"}], "策划": [{"subject": "北
京大学图书馆", "object": "阅读文化节"}, {"subject": "北京大
学图书馆", "object": "书香燕园"}, {"subject": "北京大学图书
馆", "object": "主题讲座系列"}, {"subject": "北京大学图书馆
, "object": "文化体验系列活动"}, {"subject": "北京大学图书
馆", "object": "中华文明知识竞赛"}, {"subject": "北京大学图
书馆", "object": "图书漂流活动"}, {"subject": "北京大学图书
馆", "object": "通识书专题书展"}], "应用": [{"subject": "线
装书装订体验", "object": "文化体验系列活动"}, {"subject": "学
生读书沙龙", "object": "书香燕园"}], "展示": [{"subject": "书
展", "object": "中国文学经典"}, {"subject": "文化展", "ob-

ject": "中国文学经典"}, {"subject": "文化艺术体验", "ob
ject": "中国文学经典"}], "依托": [{"subject": "中华文明知
识竞赛", "object": "中华文明史"}, {"subject": "阅读文化节",
"object": "中华文明史"}], "吸引": [], "支撑": [{"subject": "中
华文明史", "object": "中华文明知识竞赛"}, {"subject": "中华
文明史", "object": "阅读文化节"}], "参与": []]

3.4 知识图谱的可视化呈现

利用提取的相关实体及关系集合,进行实体融合与消歧
并去除含义相同的关系对后,为可视化知识图谱的直观简
洁,对数量排名前30的实体的利用Neo4j软件绘制高校图书
馆传统文化阅读推广的知识图谱,如图8所示。以相关实体
为节点,橙色、蓝色、红色、粉红色与绿色分别代表单位机构、
活动形式、文化内容、推广策略以及受众群体类别的实体,实
体间的关系名也均在节点间的连线上。

图8中,“高校图书馆”作为核心节点,连接与传统文化
阅读推广相关的多个实体,例如,“高校图书馆”与“文化展览
主题书展文化讲座”等活动形式直接相关,同时也与“主题阅
读月”“数字资源推广线上线下”等推广策略紧密相连,展现
出高校图书馆作为文化传承的主要阵地,如何通过组织丰富
多样的活动与推广策略来发扬传统文化。

此外,图谱进一步揭示高校图书馆在传统文化推广过程
中,通过多种路径和策略实现文化传播与推广目标的多维联
系,涵盖各类文化活动、推广策略和受众群体,体现出高校图
书馆作为文化传递与教育平台在现代社会中的多元作用。
在图8中,“高校图书馆->文化讲座”以及“文化讲座->线上

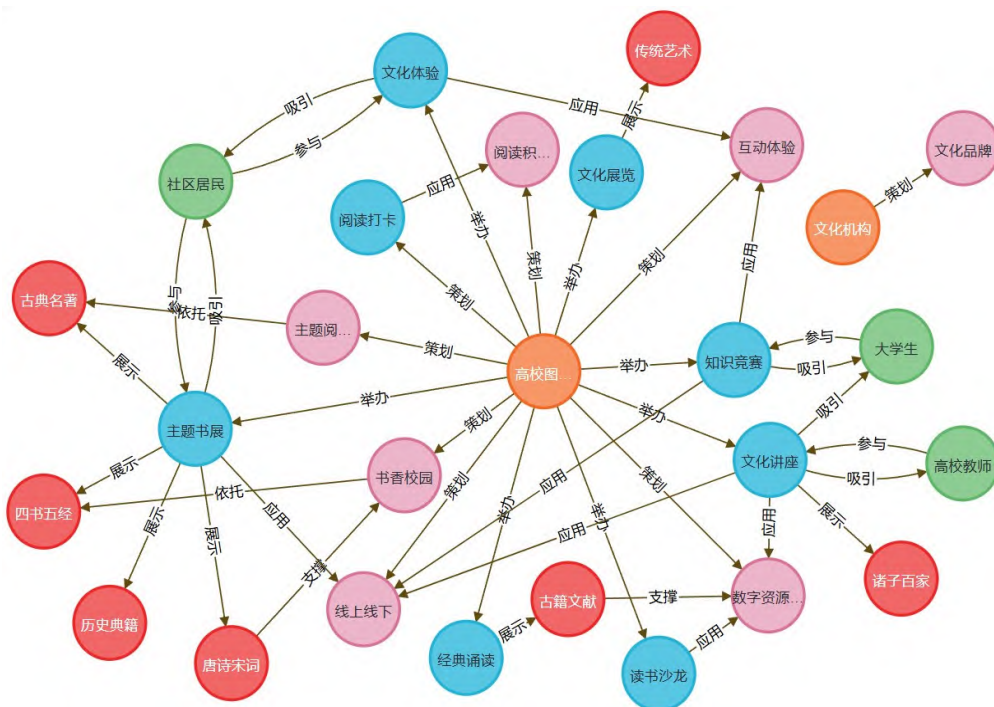


图8 高校图书馆传统文化阅读推广知识图谱

Figure 8 Knowledge graph for traditional culture reading promotion in higher education libraries

线下”路径,突显高校图书馆通过讲座等学术活动形式,通过线上线下的推广,将传统文化知识策略有效传递给受众群体的具体实践;同时,“高校图书馆->主题书展”路径展示高校图书馆在推动文化传播过程中,如何通过主题书展这一活动形式展示传统文化内容,以增强受众群体对传统文化认同感与教育影响力。

进一步分析“文化讲座->大学生”“知识竞赛->大学生”路径,展现高校图书馆在高等教育中作为文化推广平台的重要作用,高校图书馆通过组织多样化的文化活动,像是“经典诵读”“读书沙龙”与“阅读打卡”等活动形式,吸引大学生群体积极参与传统文化的传承与推广活动,拓展传统文化认同和价值认知。此外,知识图谱中“数字化技术”等技术手段路径的展现,表明现代技术在传统文化传播中的实践应用,不仅能提高文化内容的传播效率,也能进一步拓宽其受众基础,使传统文化在信息化、数字化背景下得以更广泛地传播与认同^[37]。

4 传统文化阅读推广知识图谱的创新应用

借助知识图谱,高校图书馆能够有效整合传统文化资源,创新传统文化阅读推广的路径。首先,利用知识图谱辅助传统文化阅读推广案例库的建设,通过积累和分析案例,为高校图书馆提供有效的推广策略与活动设计^[38];其次,构建智能化问答系统,以提升传统文化推广的互动性与精准性,帮助受众快速获取相关需求信息;最后,分析挖掘传统文化阅读推广路径,优化文化传播策略,提升推广效果。通过传统文化阅读推广知识图谱创新应用的实施,能够为高校图书馆提供数字化转型的支持,充分发挥高校图书馆文化传承的阵地作用。

4.1 辅助传统文化阅读推广案例库建设

高校图书馆利用现代技术能够推动传统文化的普及和传承,知识图谱的构建为高校图书馆提供了一个全新的工具,以帮助建立一个传统文化阅读推广案例库。对知识图谱的分析与应用,可以有效整理、管理并挖掘出以往和当前典型的文化推广案例,为日后的文化推广活动提供借鉴与参考。传统文化阅读推广案例库将不仅是一个纯粹的存储系统,更是一个需要动态更新、智能化运营的推广平台。

通过构建知识图谱,图书馆能够整合并可视化传统文化推广活动的多个层次,包括但不限于“活动形式文化内容”“受众群体”和“推广策略”等不同层次的元素,通过梳理能够协助高校图书馆在活动设计、策略选择和受众分析等方面作出更加精准的决策。例如,图谱中的“高校图书馆->文化讲座”与“文化活动->线上线下”路径,展示活动形式以及推广策略的具体应用。

在传统文化阅读推广案例库的建设过程中,利用知识图谱对实体和关系的归纳与分析,能够为不同高校图书馆的文

化活动设计提供借鉴。例如,通过分析图谱中“高校图书馆->主题书展”路径,图书馆可以总结出成功的书展活动类型、受众需求和推广策略。进一步分析“文化活动->经典诵读推广策略->数字化技术”路径,通过创新的数字化手段结合经典文化形式进行有效传播。知识图谱的应用不仅辅助传统文化阅读推广案例库建设,还增强传统文化活动的个性化和精准化,使文化传播更加符合受众的实际需求和时代特色。

4.2 构建智能化问答系统

知识图谱的语义关联和结构化特性为高校图书馆构建智能化问答系统提供技术支撑,通过与知识图谱的深度融合,智能问答系统能够准确解析用户对传统文化资源的查询意图,并基于图谱中的实体关系快速定位答案。例如,当用户询问“《诗经》中的经典篇目有哪些?”时,系统不仅能列出经典篇目名称,还能自动关联相关的注释、研究文献、讲座视频等资源,甚至能够推荐相关的线下活动。通过这种方式,系统不仅仅提供简单的查询答案,而是多层次、全方位呈现与传统文化相关的内容,从而优化传统文化阅读推广路径,促进文化的多元传播。

智能化问答系统需要具备动态知识更新的能力,通过持续抓取学术数据库、文化机构开放数据等外部资源,系统能够不断扩展和更新知识图谱,确保问答内容的时效性和权威性。以“文化讲座”作为例子,系统能够根据图谱中“高校图书馆->文化讲座”路径实时更新最新的讲座信息和讲座视频。此外,系统还可以通过图书馆的微信公众号、APP等平台提供“即问即答”服务,为用户提供便捷的信息查询入口,并积累用户的行为数据,帮助图书馆优化资源配置和推广策略,提高文化活动的参与度和影响力。

4.3 挖掘传统文化阅读推广路径

知识图谱的动态关联与网络化特征为高校图书馆挖掘传统文化阅读推广路径提供了数据驱动的分析路径。通过整合活动形式、受众行为、推广策略等多维度数据,知识图谱能够构建“活动-策略-效果”的关联网络,并利用路径挖掘算法识别高价值传播链路。例如,通过分析图8所示的知识图谱,能够发现“高校图书馆->文化讲座->大学生”这一路径的传播效率显著高于其他路径,此路径表明学术性活动结合混合式推广模式,即线上线下结合,对青年群体尤其是大学生群体,更具吸引力。这一结果不仅揭示文化活动设计的潜在规律,还能通过量化指标,如参与度、资源转化率,评估不同路径的实际效果,为高校图书馆传统文化阅读推广的策略迭代提供科学依据。通过这样的路径分析,图书馆能够优化推广策略,确保文化资源能够以最高效的方式传递。

参考文献

- [1] 龙春芳,董梦华.高校图书馆传统文化经典阅读推广创新

- 探索——以“百部经典校园行”活动为例[J].图书馆工作与
研究,2024(4):84-90.
- [2] 刘梦佳.新时代高校图书馆中华优秀传统文化阅读推广
的问题与对策[J].教育理论与实践,2024,44(27):18-
22.
- [3] 邓颖.文化自信视域下高校图书馆中华优秀传统文化阅
读推广策略研究[J].绥化学院学报,2023,43(2):136-
138.
- [4] 洪佳惠.一种效果历史:人文视角下的非结构化档案数据
知识化研究[J].山西档案,2023(1):78-85.
- [5] 张玲玲,黄务兰.基于ChatGPT API和提示词工程的专
利知识图谱构建[J].情报杂志,2025,44(3):180-187.
- [6] 许剑颖,冯桂珍.ChatGPT赋能档案服务:技术特征、应
用场景与实现路径[J].山西档案,2023(6):111-120.
- [7] 于梦文,崔浩男.基于创新扩散理论的高校校史馆知识服
务优化研究[J].山西档案,2023(1):63-68.
- [8] 刘梦佳.新时代高校图书馆中华优秀传统文化阅读推广
的问题与对策[J].教育理论与实践,2024,44(27):18-
22.
- [9] 穆允军,于天指,许文迪.文化传承视域下大学图书馆传
统文化阅读推广研究——以山东大学图书馆为例[J].图
书情报工作,2024,68(5):32-41.
- [10] 崔钰琪,杨红英.全民阅读视域下民族传统体育文化数
字阅读推广研究[J].科技与出版,2022(7):90-94.
- [11] 黄运红,朱婧雅,于静,等.基于文化符号体系的“激发—
推动型”双主体互动模式探究——以传统文化阅读推广
为案例[J].大学图书馆学报,2023,41(2):87-94.
- [12] 汪全莉,邹雅婷,方诗.图书馆青少年传统文化阅读推广
活动生命力评估及提升探赜[J].图书馆建设,2024(2):
163-173.
- [13] 陈筱琳.公共文化服务与流动儿童成长关系研究——兼
论面向流动儿童的传统阅读推广[J].图书馆建设,
2021(1):157-166.
- [14] 孙媛媛,张玲,于静,等.高校图书馆中华传统文化立体阅
读推广探究——以北京师范大学图书馆“二十四节气”
阅读推广为例[J].图书情报工作,2020,64(9):57-64.
- [15] 伍力.公共图书馆中华优秀传统文化阅读推广体系构建
研究[J].图书馆建设,2020,(S1):130-131,135.
- [16] 黄俏娟,曹存根,王亚,等.基于大语言模型的事件常识知
识图谱扩展方法[J/OL].软件学报,1-34[2025-04-29].
<https://doi.org/10.13328/j.cnki.jos.007262>.
- [17] 冯钧,畅阳红,陆佳民,等.基于大语言模型的水工程调度
知识图谱的构建与应用[J].计算机科学与探索,2024,18
(6):1637-1647.
- [18] 陈娟,赵新潮,隋京言,等.故事启发大语言模型的时序知
识图谱预测[J].模式识别与人工智能,2024,37(8):715-
728.
- [19] 杨燕,叶枫,许栋,等.融合大语言模型和提示学习的数字
孪生水利知识图谱构建[J].计算机应用,2025,45(3):785-
793.
- [20] 司悦航,成清,黄金才,等.基于大语言模型的时序知识图
谱推理模型蒸馏方法[J].指挥与控制学报,2024,10(6):
712-719.
- [21] 张琪,钟昊.大语言模型驱动的知识图谱实体摘要的次
模优化方法[J].计算机科学与探索,2024,18(7):1806-
1813.
- [22] 张鹤译,王鑫,韩立帆,等.大语言模型融合知识图谱的问
答系统研究[J].计算机科学与探索,2023,17(10):2377-
2388.
- [23] 张学飞,张丽萍,闫盛,等.知识图谱与大语言模型协同的
个性化学习推荐[J].计算机应用,2025,45(3):773-784.
- [24] 周春莲.人工智能技术在高校档案管理中的比较优势及
实践应用[J].山西档案,2023(1):126-128.
- [25] 陈雨,陈海玉,李文惠.数字人文视域下东巴历史档案资
源知识聚合与服务研究[J].山西档案,2023(4):100-109.
- [26] GROOTENDORST M.BERTopic:Neural topic model-
ing with a class-based TF-IDF procedure[J].Computer
Speech and language,2022(1):1109-1205.
- [27] VASWANI A, SHAZEER N, PARMAR N, et al. Atten-
tion Is All You Need[C]//Proceedings of the 31st Interna-
tional Conference on Neural Information Processing Sys-
tems.Red Hook: Curran Associates,2017:6000-6010.
- [28] REIMERS N, GUREVYCH I. Sentence-BERT: sen-
tence embeddings using siamese BERT-networks[C]//
2019 Conference on Empirical Methods in Natural Lan-
guage Processing and the 9TH International Joint Con-
ference on Natural Language Processing,2019.
- [29] MCINNES L, HEALY J, MELVILLE E. UMAP: Uniform
manifold approximation and projection for dimension re-
duction[J].Journal of Computational Biology,2018,20(7):
738-750.
- [30] MCINNES L, HEALY J, ASTELS S. hdbSCAN: Hierarchical
density based clustering[J]. The Journal of Open Source
Software,2017,2(11):205-209.
- [31] CARBINELL J, GODLSTEIN J. The use of MMR,
diversity-based reranking for reordering documents and
producing summaries[J].ACM SIGIR Forum,2017,51(2):
209-210.
- [32] GUO D, YANG D, ZHANG H, et al. Deepseek-r1: In-
centivizing reasoning capability in llms via reinforcement
learning[J].Computer Speech and language,2025(1): 110-
121.
- [33] 洪佳惠.技术、经验与民主化——知识图谱应用于档案
资源开发的认识论研究[J].山西档案,2023(1):86-92,41.

- [34] 吕树月.基于情感分类和主题挖掘的冬奥微博舆情演化研究[D].桂林:桂林电子科技大学,2023.
- [35] WEI J, WANG X, SCHUURMANS D, et al. Chain-of-thought prompting elicits reasoning in large language models[J]. *Advances in Neural Information Processing Systems*, 2022, 35: 24824-24837.
- [36] 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[2025-03-17].
- [37] 胡慧慧, 赵雪芹. 基于本体的桥梁文化遗产档案知识图谱构建研究[J]. *山西档案*, 2023(6): 42-57.
- [38] 张玲怡, 王向女. 我国档案馆网站在线知识开发与服务研究[J]. *山西档案*, 2022(4): 104-111.

(责任编辑:张连峰)

Construction of Knowledge Graph and Optimization Path of Traditional Culture Reading Promotion in College Libraries Based on Big Language Modeling

WANG Lili

(Library, Jilin University, Changchun 130012, China)

Abstract: 【Purpose/significance】 In the context of the new era, to explore the knowledge graph construction process of traditional culture reading promotion in college libraries based on the big language model, to optimize the reading promotion path through the visual graph, to promote the deep excavation and systematic organization of traditional culture resources, and to innovate the reading promotion mode in college libraries. 【Method/process】 This study proposes a knowledge graph construction process of traditional culture reading promotion content based on large language model. After collecting data, we first carry out data cleaning and processing, use BERTopic topic model to extract entity categories and relationship labels, and then use deepseek-r1 large language model for entity extraction and relationship recognition, and finally import the entity relationship triples into Neo4j graph database, which can be used to optimize the reading promotion path through visual graph. Finally, the entity-relationship triples are imported into the Neo4j graph database to achieve the visualization of the knowledge graph. In addition, we use the specific content of Baidu search and WeChat search platform to construct the knowledge map of traditional culture reading promotion in college libraries, and explore its innovative path mode in traditional culture reading promotion. 【Result/conclusion】 The results of the study show that the knowledge mapping method based on the deepseek-r1 model can clearly present the organization, activities and promotion strategies in traditional culture reading promotion, forming a structured knowledge expression system; further analysis of the knowledge content of traditional culture reading promotion can assist in the construction of the traditional culture reading case base, help build an intelligent Q&A system, assist in the mining of traditional culture reading promotion paths, and assist in the construction of the traditional culture reading case base. mining traditional culture reading promotion paths, in order to play the role of the position of cultural heritage in college libraries. 【Innovation/limitation】 A knowledge graph construction method of traditional culture reading promotion content based on big language model is proposed, and the knowledge elements are mined for traditional culture promotion in college libraries, but due to the limitation of computer resources, the big model interface of the third-party platform can be used in the future research.

Keywords: big language modeling; BERTopic; knowledge graph; traditional culture; reading promotion