

<https://doi.org/10.3799/dqkx.2025.127>



基于知识图谱检索增强生成的滑坡监测预警系统

吴润泽^{1,2}, 李浩³, 梅红波³, 王成彬³, 朱敏毅¹, 王红群¹, 张亮⁴, 胡光鸿⁴,
马明杰⁴, 望致文⁴, 张隆隆³, 黄旻³, 李振华^{4,5*}

1. 中国地质调查局武汉地质调查中心, 湖北武汉 430205
2. 高坝大库运行安全湖北省重点实验室(中国长江三峡集团有限公司), 湖北宜昌 443133
3. 中国地质大学(武汉)资源学院, 湖北武汉 430074
4. 中国地质大学(武汉)计算机学院, 湖北武汉 430078
5. 新疆理工学院信息工程学院, 新疆阿克苏 843100

摘要: 针对当前滑坡防治工作中文字报告中的知识提取和利用问题, 探索出一套基于地灾资料和数据的知识图谱构建和运用方法, 促进地灾防控工作从数据管理到知识管理转型. 对 182 份报告、22 篇规范、5 篇文献、4 668 个滑坡点和 313 项治理工程的基础数据、3 年来的历史灾情数据、当前滑坡监测数据和预报的天气数据进行了知识提取和图谱构建, 并采用检索增强生成技术实现多源知识的整合和输出. 建成了一个具有 12 797 个实体、34 873 个关系、9 658 个 chunk 的知识图谱系统, 实现了 4 518 个滑坡的 10 天预警. 知识图谱技术可有效提取文字报告中的知识, 检索增强生成技术可有效整合多个知识来源, 提高知识问答的准确性.

关键词: 三峡库区; 滑坡监测预警; 知识图谱; 大语言模型; 检索增强生成; 工程地质学.

中图分类号: P694

文章编号: 1000-2383(2025)10-4125-12

收稿日期: 2025-03-04

A Landslide Monitoring and Early Warning System with Retrieval-Augmented Generation Enhanced by Knowledge Graph

Wu Runze^{1,2}, Li Hao³, Mei Hongbo³, Wang Chengbin³, Zhu Minyi¹, Wang Hongqun¹, Zhang Liang⁴,
Hu Guanghong⁴, Ma Mingjie⁴, Wang Zhiwen⁴, Zhang Longlong³, Huang Min³, Li Zhenhua^{4,5*}

1. Wuhan Center of China Geological Survey, Wuhan 430205, China
2. Hubei Key Laboratory of Operation Safety of High Dam and Large Reservoir (China Three Gorges Corporation), Yichang 443133, China
3. School of Earth Resources, China University of Geosciences, Wuhan 430074, China
4. School of Computer, China University of Geosciences, Wuhan 430078, China
5. School of Information Engineering, Xinjiang Institute of Technology, Akesu 843100, China

基金项目: 中国地质调查局武汉地质调查中心三峡库区后续地质灾害防治信息系统建设项目(No.0001212012AC50001).

作者简介: 吴润泽(1979—), 男, 高级工程师, 长期从事地质灾害防治信息化研究. ORCID:0000-0003-1063-1431. E-mail: jcywurunze@mail.cgs.gov.cn

*** 通讯作者:** 李振华(1969—), 男, 教授, ORCID:0000-0003-1497-4407. E-mail: zhli@cug.edu.cn

引用格式: 吴润泽, 李浩, 梅红波, 王成彬, 朱敏毅, 王红群, 张亮, 胡光鸿, 马明杰, 望致文, 张隆隆, 黄旻, 李振华, 2025. 基于知识图谱检索增强生成的滑坡监测预警系统. 地球科学, 50(10):4125-4136.

Citation: Wu Runze, Li Hao, Mei Hongbo, Wang Chengbin, Zhu Minyi, Wang Hongqun, Zhang Liang, Hu Guanghong, Ma Mingjie, Wang Zhiwen, Zhang Longlong, Huang Min, Li Zhenhua, 2025. A Landslide Monitoring and Early Warning System with Retrieval-Augmented Generation Enhanced by Knowledge Graph. *Earth Science*, 50(10):4125-4136.

Abstract: Aiming at the current problem of knowledge extraction and utilization from text reports in landslide prevention and control work, this study explores a set of methods for constructing and applying a knowledge graph based on geological disaster data and materials, promoting the transformation of geological disaster prevention from data management to knowledge management. Knowledge extraction and knowledge graph construction were conducted on basic data from 182 reports, 22 specifications, 5 literatures, 4 668 landslide points, and 313 governance projects, as well as three years of historical disaster data, current landslide monitoring data, and forecasted weather data. Retrieval-augmented generation (RAG) technology was adopted to integrate and output multi-source knowledge. A knowledge graph system with 12 797 entities, 34 873 relationships, and 9 658 chunks was established, enabling 10-day early warning for 4 518 landslides. Knowledge graph technology effectively extracts knowledge from text reports, and RAG technology efficiently integrates multiple knowledge sources, improving the accuracy of knowledge question-answering.

Key words: Three Gorges reservoir area; landslide monitoring and early warning; knowledge graph; large language model; retrieval-augmented generation; engineering geology.

0 引言

三峡库区蓄水后,受复杂地质构造与多重自然—人为因素叠加影响,该区地质灾害呈高发态势,严重威胁着该区的生命财产安全、工程运行状态及生态环境稳定.为有效应对地质灾害威胁,近二十多年来,国家在该区域开展了广泛的地质灾害调查、分析和研究工作,取得了丰硕的成果,为库区的地质灾害防治工作提供了有力支撑.然而,这些研究成果大多以研究报告、专著等形式存放,其中蕴含的丰富信息和知识难以被信息系统直接利用.新兴的知识图谱技术能够挖掘文档中的知识,同时抽取数据库中的知识,从而将原本分散在文字报告中的信息和信息系统产生的数据融合成知识网络,构建起结构化的知识体系.这不仅实现了地质灾害知识的显性化,有助于深化对地质灾害规律的认识,更重要的是,通过知识计算,能够提高地质灾害防治工作的智能化水平和及时性,进一步提升地质灾害防治的效率和效果.

知识图谱的概念由 Google 公司于 2012 年正式提出,其初衷旨在整合多个来源的信息,提升信息搜索质量.它是一种基于图的数据结构,由节点(代表实体、概念或属性)和边(代表实体之间的关系、实体与属性之间的关联)组成,以结构化的方式表示和存储知识,从而支持计算机对知识的理解、处理和应用(Singhal, 2012; Zou, 2020; Hogan *et al.*, 2021).近年来,知识图谱在智能搜索、推荐系统和自然语言处理等领域得到广泛应用(徐增林等, 2016; 漆桂林等, 2017; Abu-Salih, 2021),在各个专业领域中也展现出知识管理与应用的巨大潜力,如中医、电力、金融、航空、水利等领域(阮彤等,

2016; 冯钧等, 2019; 李新鹏等, 2019; 吕华揆等, 2020; 曹现刚等, 2021; 韩涛等, 2021; 蒲天骄等, 2021; 周逸凡等, 2024).

地球科学领域也开展了知识图谱相关研究.早期的研究偏重于知识网络的建立, SWEET 项目(Semantic Web for Earth and Environmental Terminology; Raskin and Pan, 2005)和 GeoLink 项目(Cheatham *et al.*, 2018)是地学知识图谱发展中的两个典型项目. SWEET 项目由美国国家科学基金会(NSF)资助,旨在建立一个包含定义、关系和属性信息的地球和环境科学术语库;而 GeoLink 项目则利用语义技术和链接数据的方法,整合来自不同数据源和学科领域的地学信息,构建一个跨学科、跨领域的地学知识图谱,使研究人员能更轻松地获取和理解地球科学领域的知识.近年来的研究偏重于地学数据的整合与分析, Ma *et al.* (2020)提出了一种新的深时知识图谱结构,用于更好地表示和追踪地学概念的版本信息;周成虎等(2021)系统性探讨了地学大数据背景下的知识图谱研究方向,着重阐释其在处理海量地学数据方面的技术优势; Zhang *et al.* (2022)通过 GeoKG 知识图谱的构建实践,揭示了地学知识图谱在数据完备性提升和时空关联性强化方面面临的关键挑战; Zhu *et al.* (2023)提出具有时空自适应特征的地学知识图谱框架,有效整合了多维时空特征与复杂关系表达.

具体到滑坡灾害领域,一些学者也进行了知识图谱应用方面的探索. Wang *et al.* (2018)利用自然语言处理技术从地质文献中提取关键信息,这是构建地学知识图谱的基础工作,其研究帮助解析复杂的地质数据; Fan *et al.* (2020)提出了一种结合多分支 BiGRU 和 CRF 的深度学习模型,用于从地质灾

害文献中自动识别位置、方法和数据等命名实体,并构建知识图谱;杜志强等(2020)在自然灾害应急领域提出了一种新的知识图谱构建方法,并以洪涝灾害应急知识图谱为例进行了实验验证;刘文聪(2022)构建了滑坡知识体系,设计了滑坡知识问答系统,为决策者提供知识查询,辅助决策;许强等(2023)通过自顶向下与自底向上结合的方式,构建了包括滑坡野外调查、滑坡评价等 10 大类知识为基础的滑坡知识图谱体系,通过多源异构数据融合与 Neo4j 可视化存储,实现了滑坡知识的语义关联与智能推理;邱芹军等(2023)根据地质灾害的不同阶段进行实体建模,增强了防灾减灾的系统性和针对性.Chen *et al.*(2024)提出基于知识图谱的滑坡预测方法,将滑坡预测任务转化为链接预测任务,并训练知识图谱嵌入模型来实现滑坡预测。

综上所述,知识图谱作为一项新兴技术,已在多个领域展现出其数据整合、知识提取和知识应用能力,但具体到地质灾害领域,这方面研究方向比较分散,针对单个方法或单个应用偏多,即便存在综合性的研究,也是偏理论较多,迄今为止,在地灾监测预警方面尚未发现可供日常工作的应用案例。

本研究将充分借鉴其他领域成熟的知识图谱技术和现有地灾领域的研究成果,结合三峡库区的地灾数据资料实际和监测预警要求,针对滑坡监测预警本体构建、非结构化数据的知识抽取、动态数据与静态数据融合、多源知识集成输出等难题,进行知识抽取方法上的集成创新、滑坡监测预警本体构建的理论创新和服务三峡库区工作实际的应用创新,形成一套涵盖数据和资料管理、信息抽取、知识图谱构建、知识计算和知识输出的方法体系,最终实现一个基于知识图谱检索增强生成的三峡库区滑坡监测预警系统。

1 数据来源

本研究所使用的数据主要来自于三峡库区的地质灾害监测数据、历史灾害记录、滑坡勘察报告以及治理工程数据。具体数据来源包括:

(1)结构化数据:①滑坡基本信息表,包括滑坡 4 668 个和崩塌 446 个,该表详细记录了滑坡和崩塌的相关情况,如地理位置、滑坡时间、滑坡类型、地层岩性、构造部位、规模等级等;②三峡库区治理工程数据表,该表记录了三峡库区 313 个灾害体的治理工程相关信息,包括工程名称、工程级别、工程地

理位置。通过结构化数据表,可以获取到滑坡灾害体的重要属性和构建知识图谱的关键实体,且关系型数据表天然包含属性之间的关系,为构建知识图谱提供了良好的支持。

(2)非结构化数据:①三峡库区灾害应急调查报告,包含巴东县、秭归县、巫山县等地区 2003—2020 年的 66 个报告,报告详细记录了灾害发生的时间、地点、规模等信息;②治理工程勘察报告:包含兴山、秭归、巴东等地区的共 50 个报告,主要内容是灾害体详细勘察知识及工程防治措施;③地质灾害防治信息化名词术语字典:包含地质灾害领域名词术语字典及其释义;④地质灾害防治相关标准及规范化文档:包括地质调查相关工作指南、技术要求、评估规范等 22 篇文档,如工程地质调查工作指南、地质灾害风险调查评价技术要求等;⑤滑坡领域相关书籍、科研论文共 5 篇:包括滑坡领域相关理论介绍以及案例分析,如《三峡库区滑坡监测预警理论与实践》等。

(3)监测数据:主要包括了兴山县、秭归县的监测月报(2023—2024 年共 66 份),数据内容为灾害体变形威胁对象、监测分析结果、防治建议。为了确保数据质量与可用性,对所有数据进行了预处理,包括数据清洗、去重、标准化处理等,确保数据能够准确反映滑坡灾害的发生与发展规律。

2 三峡库区滑坡监测预警知识图谱构建

图 1 为知识图谱构建流程图。通过分析滑坡的结构特征、动力成因等要素,构建了滑坡本体模型,然后收集并整理了所有三峡库区滑坡相关的结构化数据和非结构化数据,经预处理形成本研究区的语料库,最后通过知识抽取、融合、消歧等工作构建出滑坡知识图谱。

2.1 本体模型设计

(1)三峡库区滑坡监测预警本体模型设计如图 2 所示。该滑坡灾害监测预警本体模型通过不同实体之间的关联实现滑坡灾害的实时预警。其中,监测数据是系统的基础,它提供了有关滑坡体的实时信息;监测数据通过与预警模型和判据结合进行计算,生成预警结果;预警判据则来源于相关调查报告,并通过发生过的滑坡事件对其进行校正。在此过程中,滑坡体作为核心对象,与事件紧密相关,后者提供了发生时间和地点等详细信息;滑坡体的位

三峡库区滑坡地质灾害防治知识图谱构建与应用

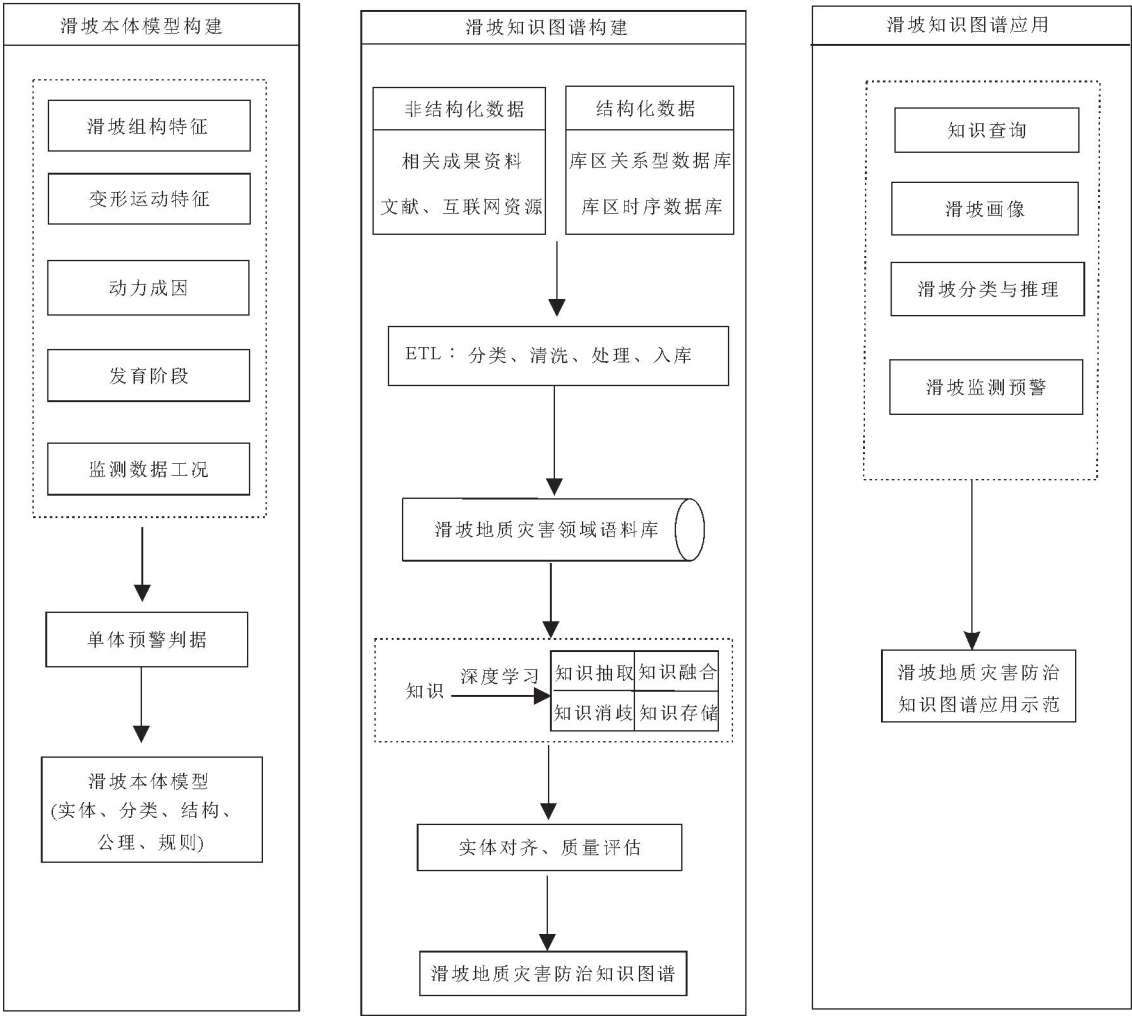


图 1 三峡库区滑坡监测预警知识图谱构建技术路线

Fig.1 Construction approach of a knowledge graph for landslide monitoring and early warning in the Three Gorges reservoir area

置和属性进一步细化了灾害的特征,并能通过类别对灾害体进行分类统计.而各类文字报告中的关键部分以 chunk 的形式存储到向量数据库中,支持不同类型滑坡及其成因理论方面的知识检索.

基于此本体模型,可完成如下功能:

- ①滑坡画像:通过灾害体及关联边,多维度刻画滑坡特征;
- ②分类统计:按类别对灾害体进行分类统计;
- ③单体预警:结合监测数据、判据及模型信息,实现滑坡单体动态预警;
- ④知识问答:利用实体与知识单元 chunks 连接,支持自然语言交互查询;
- ⑤知识推理:基于知识图谱,通过节点相似性计算挖掘潜在知识关联.

根据此本体模型,设计了滑坡监测预警知识图

谱模式层,如图 3 所示.该模式层系统地描述了滑坡预警领域的关键实体、属性及其复杂关系,主要包括临滑判据、灾情预警、监测模型、复活判据、历史滑坡和工况等核心要素,构成了滑坡风险分析与预警的知识框架.模型通过动态修正与补充机制,将监测数据与工况条件(如库水工况、降雨工况、位移数据)相结合,以稳定性系数为核心指标,综合历史滑坡数据和复活分析结果,建立了多模型预警的滑坡预警预报本体模型.

该模型依托实时监测数据接口,同步获取库水位、降雨量及位移等动态参数,结合地质静态特征与历史案例库进行时空关联分析,从而实现了静态与动态数据有机整合,通过并行计算三大模型的预警结果并依据最危原则生成最终预警等级,同时动态更新知识图谱节点以存储预警信息及决策路径,

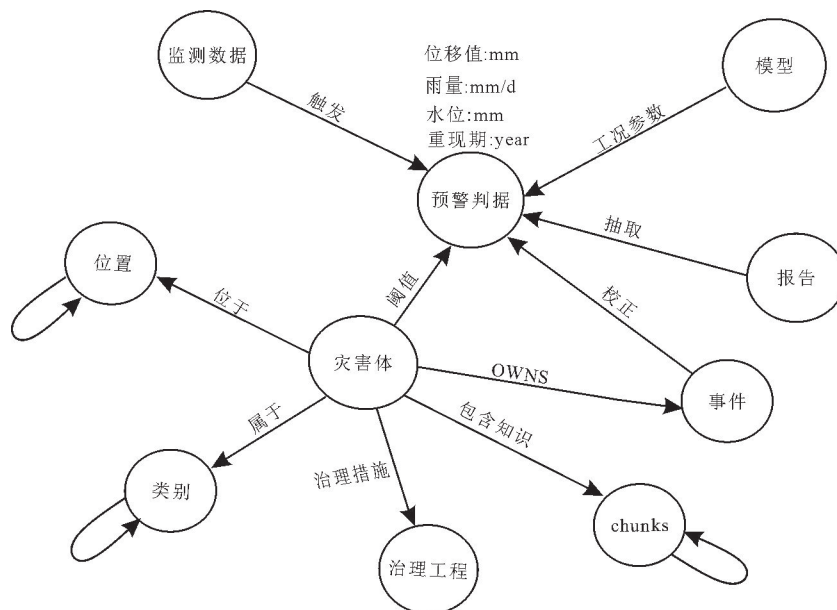


图2 三峡库区滑坡灾害监测预警本体模型

Fig.2 Ontology model for landslide monitoring and early warning in the Three Gorges reservoir area

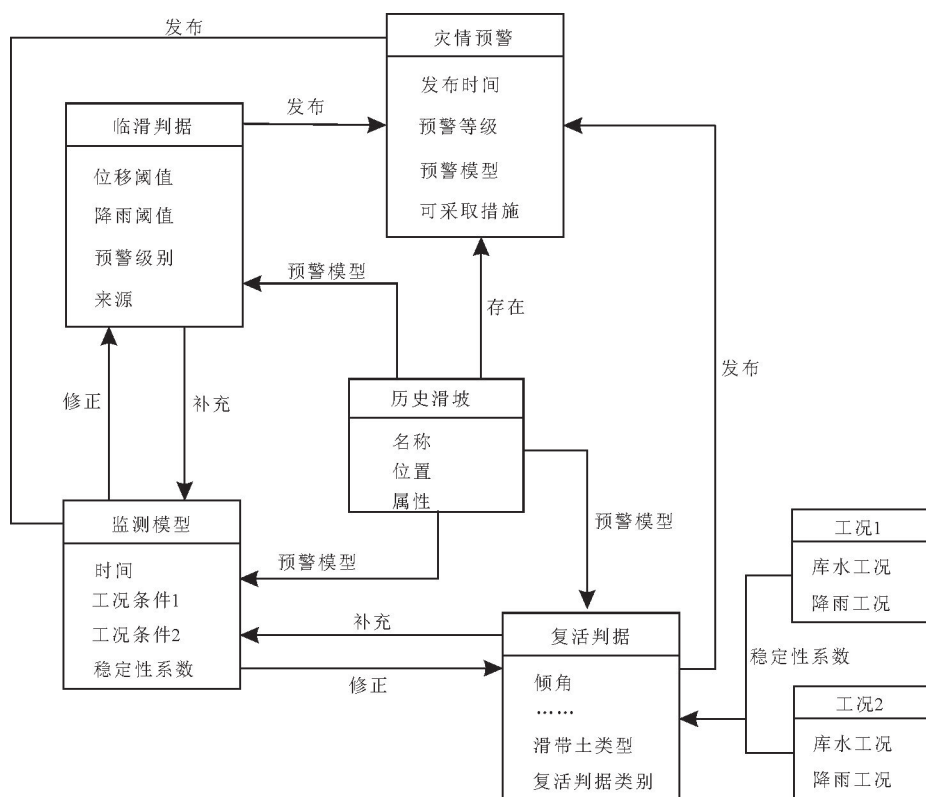


图3 三峡库区滑坡监测预警知识图谱模式层

Fig.3 Schema layer of the knowledge graph for landslide monitoring and early warning in the Three Gorges reservoir area

形成“监测—分析—预警—响应”闭环管理。

2.2 知识抽取

按数据来源,知识抽取为分结构化数据抽取和非结构化数据抽取。

2.2.1 结构化数据抽取 结构化数据表主要包括滑坡基本信息表和三峡库区治理工程信息表,涵盖了滑坡灾害体的基本属性信息,如滑坡名称、地理位置、滑坡类型、滑体岩性等内容。通过将模式层中

定义的实体类型与属性映射到 Neo4j 数据库,可以实现对滑坡相关信息的高效存储和管理,为后续的查询、分析和可视化提供支持.结构化数据导入 Neo4j 示例如图 4 所示,图中该滑坡相关属性是由抽取的结构化数据映射得来.

该映射方法主要包含 4 个核心层面:实体映射、属性映射、关系映射和数据完整性处理.在实体映射层面,将滑坡记录映射为图数据库中的“历史滑坡”主实体节点,同时建立“滑坡类型”“工程治理”“规模等级”和“县区名称”等关联实体节点.属性映射层面涵盖了标识属性(滑坡 ID)、名称属性(滑坡名称和乡镇名称)、地理属性(经纬度)和时间属性(滑坡年代)等基础属性,以及包括滑体性质、岩性、结构等地质特征属性和坡度、坡向、面积、体积等几何属性的映射转换.在关系映射层面,构建了分类关系、治理关系、等级关系和地理关系四类关联关系,形成完整的图数据模型.数据完整性处理方面,通过空值处理机制、特殊字符处理和数据类型转换确保了数据的规范性和可用性.这种多层次的映射方法不仅保持了原始数据的完整性和关联性,而且有利于后续基于图数据的分析和查询.

2.2.2 非结构化数据抽取 在本研究中,采用了 3 种方法进行非结构化数据的知识抽取:基于专家的知识标注、基于深度学习模型的知识学习和基于大语言模型的知识理解.

(1)基于 Prodigy 的知识标注.在三峡库区滑坡知识图谱模式层的指导下,使用 Prodigy 工具对非结构化文本进行手动标注.Prodigy 是一款高效的文

本标注工具,它结合了深度学习模型和基于规则的标注方法,能够实现对非结构化文本数据的快速实体标注、关系提取及语义识别.该工具提供了一个简洁而直观的界面,支持用户快速高效地进行标注任务,如图 5 所示,涂色部分标注了滑坡和地理位置这两类实体.

(2)基于 Transformer 的知识学习.针对人工标注的实体识别与关系抽取数据集,采用基于 Transformer 的深度学习模型进行双任务联合训练.实验结果表明,基于 Transformer 架构训练的模型在各实体类别的性能指标(召回率、精确率、F1 值)均稳定达到 0.85 以上,充分验证了该模型在滑坡领域信息抽取任务中的有效性.

针对预测文本中的实体标注误差,执行实体标注纠错处理后,将人工校验数据导出为 JSON 格式,通过 Python 脚本自动化转换为 Cypher 语句,最终导入 Neo4j 图数据库.如图 6 所示,展示了与张家湾滑坡相关的多种关联信息.其中,张家湾滑坡是核心节点,与之相关联的节点包括其地理位置、滑坡类型、规模等级、复活判据分类等.

(3)基于大语言模型的知识理解.在基于深度学习实现知识抽取与图谱构建的基础上,为确保知识抽取的全面性与准确性,本研究引入大型语言模型 (large language model, LLM) 优化抽取流程.LLM 凭借其强大的语义理解与生成能力,可在多维语义空间补充传统深度学习模型可能遗漏的知识节点,有效提升知识覆盖度.

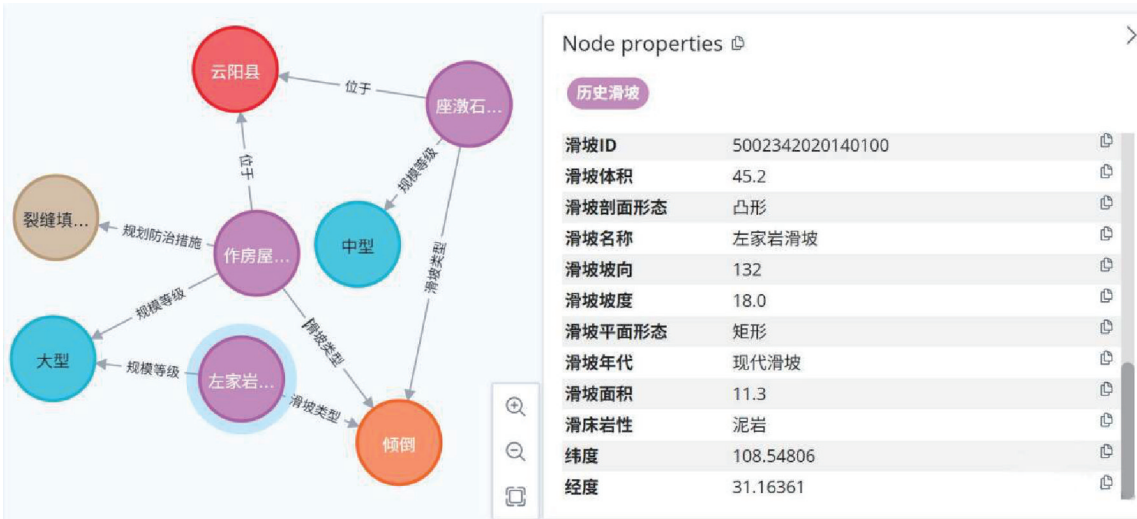


图 4 基于结构化数据映射的知识图谱

Fig.4 Knowledge graph based on structured data mapping

通过大模型接口调用,对初步构建的知识图谱实施二次验证与扩展:一方面利用其上下文语义理解能力捕捉长尾知识及隐含关系,补全初步抽取的语义缺口;另一方面通过领域知识推理增强图谱的语义一致性和逻辑严谨性.同时,LLM的动态学习特性可适配知识更新需求,实现新增数据的快速抽取与图谱迭代,显著增强知识图谱的时效性.

在模型选型上,采用 OneKE 框架 (Gui *et al.*, 2024). 该框架由蚂蚁集团与浙江大学联合研发,支持中英文双语、多领域多任务的泛化知识抽取,并提供完善的工具链支持,以开源形式贡献至 Open-KG 开放知识图谱社区.通过人工定义 Schema 解释指令,可直接对非结构化文本(如 doc/txt)进行批量实体关系抽取,生成结构化 JSON 等格式数据,比如:{"滑坡":["大岭西南滑坡"],"滑坡类型":[],"所在地区":["锣鼓洞河左岸"],"秭归县沙镇溪镇三星店三组"],"滑坡名称":[""],"监测预警级别":["]}.

实验结果表明,LLM在滑坡领域信息抽取任务中显著提升了实体抽取的完整性与关系识别的准确性,同时展现出更强的领域泛化能力.最终构建的知识图谱有效抽取了滑坡地理位置、规模、降雨阈值等核心预警判据信息及滑坡类型划分体系,尤其在复杂长短期实体构成的预警判据抽取中,LLM对多语义嵌套场景的解析优势显著.如图7所示,该

图谱通过 LLM 增强抽取,实现了结构化知识的系统性整合,形成包含多维度预警指标与实体关联关系的体系.

2.3 知识图谱构建

本研究构建的知识图谱包含 12 797 个异构节点,涵盖了历史滑坡实体、地质特征、工程措施和滑坡判据等多个领域的知识实体.图谱中包含 34 873 条多样化的关系边,包括滑坡分类关系、空间位置关系、成因关系和防治关系等,形成了一个高度互联的知识网络结构.如附图 1 为三峡库区滑坡知识

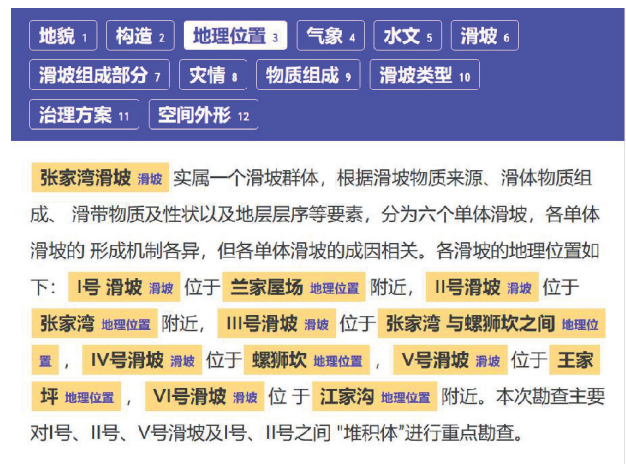


图 5 基于 Prodigy 的实体与关系标注
Fig.5 Entity and relation annotation using Prodigy

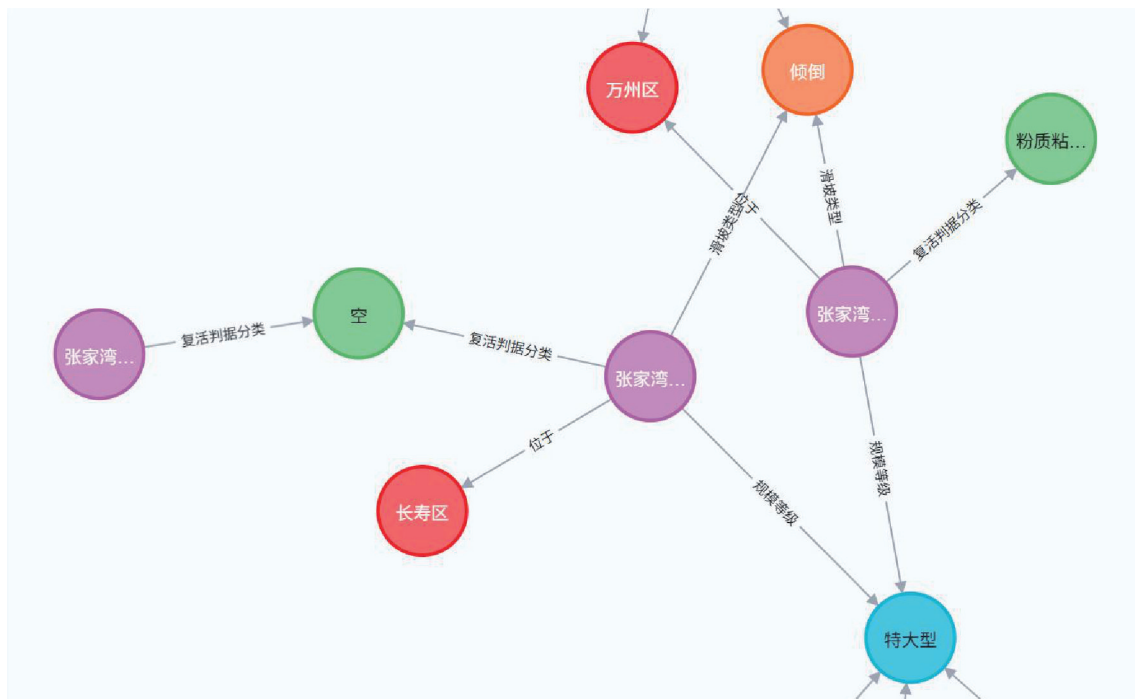


图 6 张家湾滑坡勘察报告信息抽取实例

Fig.6 An example of information extraction from Zhangjiawan landslide investigation report

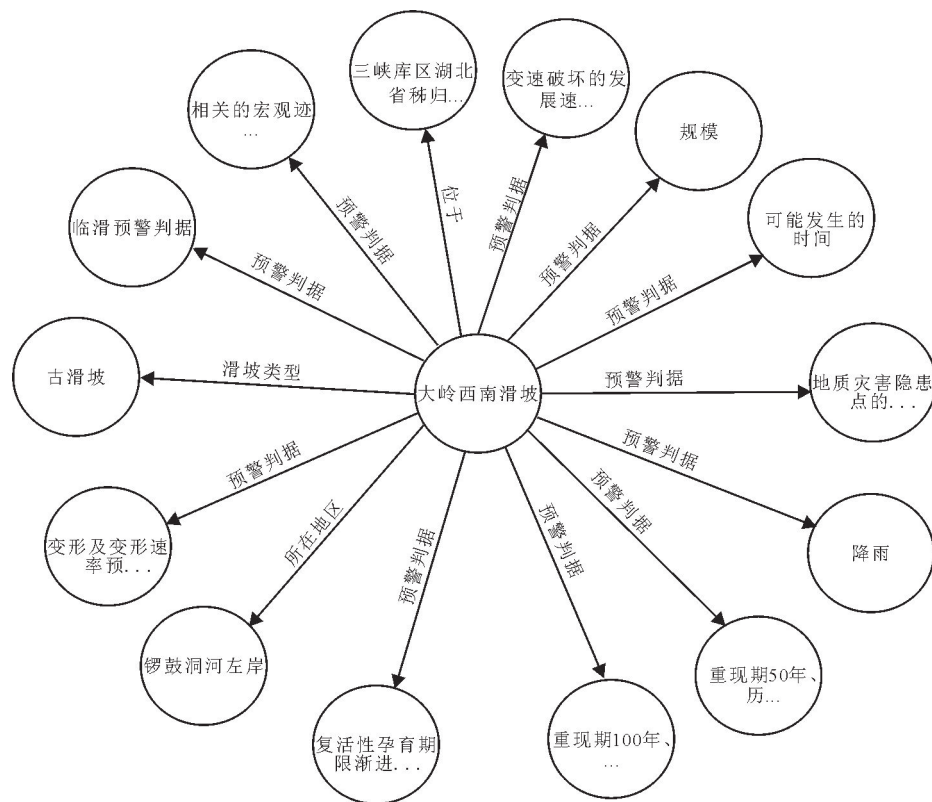


图7 基于大型语言模型的知识图谱构建实例
Fig.7 Knowledge graph construction based on large language model

图谱示意图,呈现了图谱中部分节点的核心信息:紫色节点代表滑坡实体,红色节点对应位置实体,橙色节点表示滑坡类型,棕色节点为治理措施.节点间的关系通过连线上的标签显示,呈现不同实体类型间的语义关联.

3 三峡库区滑坡监测预警知识图谱应用

三峡库区滑坡监测预警知识图谱系统的主要功能有知识管理、滑坡画像、监测预警、检索增强生成、知识问答等功能.

3.1 知识管理

附图2为知识管理模块页面.该模块旨在实现对非结构化数据的高效管理和利用.用户可以通过上传文档模块将湖北灾情数据、应急报告、相关技术规范等各类非结构化文档导入系统.这些文档经过预处理后,会被存储到向量数据库中,为后续的智能检索、相似性匹配及知识推理提供支持.模块支持多种文件格式的上传,并能根据文档内容自动提取关键信息,形成结构化索引,为灾情数据分析和应急决策提供数据支撑,显著提升知识管理的效

率和精度.

3.2 滑坡画像

本系统的知识图谱可视化基于Neo4j数据库构建,不同颜色的节点表示不同的节点类型,不同节点之间存在不同的关系.如附图3所示,显示了张家湾滑坡的知识图谱画像,展示了与该滑坡有关的33个节点和38个关系.

3.3 监测预警

由于实际边坡形状不规则,所受荷载条件复杂,且滑动面在失稳前难以确定,因此,基于刚体假设的极限平衡法,和有限元等数值模拟方法无法准确地计算边坡稳定系数(张忠孝, 2023).本研究采用传递系数法计算边坡稳定系数,该方法通常与数值分析相结合,计算在降雨和地震工况下边坡的稳定系数,为解决此种复杂工况下的稳定性计算问题提供了实现路径.

为实现滑坡灾害的精准风险评估,本研究完成了知识图谱系统与滑坡监测系统的实时数据接口集成工作.该模块从监测数据库中实时提取关键工况指标,包括当前及历史时段的位移量、降雨量和库水位数据.这些关键参数随后输入预编译的

MATLAB算法,该算法基于当前工况条件,结合滑坡稳定性分析模型计算稳定性系数,评估滑坡的实时稳定状态,并将计算结果反馈至系统。

监测预警模块集成了单体滑坡判据和滑坡稳定性计算模型,并基于实时数据动态和天气预测数据分析滑坡风险,实现了基于多模型的实时预警体系。一旦监测数据触发预警阈值,系统立即生成预警信息并将最危预警结果展示在地图上,并以不同颜色和图标区分预警等级。如附图4所示展示了当前日期(2025年6月19日)后10天预测预警情况。

3.4 检索增强生成

为提升生成内容的准确性、专业性和时效性,降低大模型直接生成答案产生的幻觉问题,本研究采用检索增强生成(RAG)技术,该技术将信息检索与大语言模型的生成能力相结合,通过优先检索知识图谱来增强模型的回答。它包含检索、增强、生成3个阶段。

在检索阶段,本研究整合了多数据源,包括三峡滑坡基本信息图数据库、滑坡报告知识片段向量数据库、灾险情数据库、滑坡预警预报数据库等。针对用户问题,系统将同时调用图数据库检索工具(Neo4j)、向量数据库检索工具(Milvus)及关系型数据库检索工具(MySQL),分别从滑坡知识图谱实体关系网络、滑坡报告等专业文档语义向量空间、结

构化历史数据中提取关联信息,形成对大语言模型(LLM)的知识补全。

增强阶段构建多源信息与大语言模型的深度融合链路,通过对检索到的滑坡知识图谱的三元组、向量文档的语义片段、关系数据的字段集合进行提示词工程优化,引导大语言模型生成响应结果,实现“事实性信息基于检索数据,逻辑性推理依托模型能力”的混合生成模式,显著提升回答的准确性与可信度。

生成阶段负责与用户交互,将增强阶段输出的结构化答案转化为自然语言文本,并以友好的界面形式反馈给用户,同时支持多轮对话和上下文理解。具体的框架如图8所示。

3.5 知识问答

为实现滑坡灾害信息的高效检索与知识推理,本研究构建了一个集成多源知识库和工具链的智能问答系统。该系统基于滑坡专家知识构建的对话框架,结合先进的自然语言处理技术,能够精准解析用户查询,并从多维数据源中获取相关信息,其回答类型包括滑坡数据查询、知识查询、统计分析、统计推理、预测预报多类问题。

如附图5所示为数据查询型问答场景,查询特定滑坡的基本信息。针对此类问题,知识图谱数据库与Milvus向量数据库的知识较为完善,可提供全

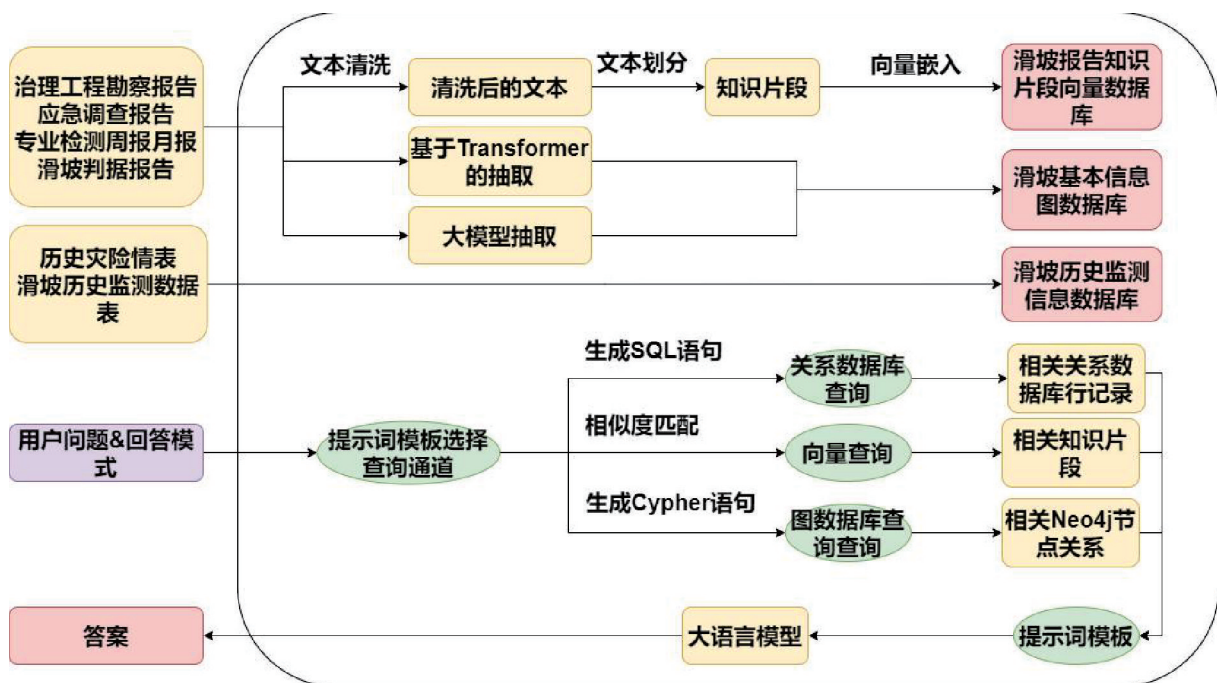


图8 三峡库区滑坡监测预警系统RAG问答架构

Fig.8 RAG Q&A architecture of the landslide monitoring and early warning in the Three Gorges reservoir area

面且精准的回答。

如附图 6 所示为知识查询型问答场景,查询动水压力型滑坡的特点。针对此类问题,Milvus 向量数据库因存储大量相关文档,可基于向量检索能力实现多维度技术内容的匹配与关联响应,并进行总结后输出。

如附图 7 所示为统计分析查询型问答场景,示例问题为“2023 年有多少灾险情?这些灾险情有什么共同特点?”。在此类问答中,系统首先提取问题中的时间关键词与灾险情关键词,定位至灾险情数据库并基于时间条件筛选数据,通过关系型数据库提取结构化统计数据解答“数量”类问题,此外大模型会将查询到的结果进行汇总回答并补充相关知识。

如附图 8 所示为统计推理问答场景,查询满足特定参数下的滑坡数目以及共同特点,在此类问答中,大模型首先提取问题中的查询目标与查询条件,随后由图数据库和关系型数据库执行统计分析任务,AI 模块则负责注入与问题背景相关的延伸知识,实现数据驱动的分析结果与领域背景信息的有机融合。

如附图 9 所示为预测预报型问答场景。针对此类问题,后台系统将基于未来数天的工况参数,通过专业模型计算滑坡稳定性系数、判据匹配等,进而对潜在高风险滑坡进行预警。

4 结论

本研究针对三峡库区滑坡监测预警的日常工作要求,构建了基于知识图谱检索增强生成的滑坡监测预警系统,为类似的地质灾害防治工作提供了参考性的解决方案。取得的主要成果如下:

(1)构建了三峡库区滑坡监测预警本体模型。面向滑坡地质灾害的孕育过程,满足地质灾害“监测—识别—预警—处置”业务需求,系统梳理了地质灾害领域的实体、语义关系和关联属性信息,构建一个“监测—分析—预警—响应”地质灾害领域本体模型。

(2)盘活了 209 本“沉睡”的资料,抽取出了相应的信息和知识。所处理的资料包括:学术论文 5 篇,规范文档 22 篇,监测月报、勘察报告等多类报告 182 份。通过知识图谱的建立,这些分散且难以利用的资料得以结构化和关联化,形成了一个涵盖多维度信息的系统知识体系,为滑坡监测预警、数据挖掘

与智能分析提供了数据基础和理论支撑。

(3)建立了三峡库区滑坡监测预警知识图谱。该知识图谱包括 12 797 个实体、34 873 个关系、9 658 个 chunk,融合了历史地质灾害数据、灾险情数据、临滑判据报告、监测预警数据、复活判据数据以及防治工程数据、以及各类资料等多源异构数据,并以结构化的形式存储。它不仅实现了地质灾害相关信息的语义化表达,还建立了数据之间的复杂关联关系,如灾害诱因、时空分布特征、风险评估指标与预警阈值等,从而为滑坡监测与预警提供了精准的数据支撑与智能化的决策依据。

(4)实现了 4 518 个滑坡的每日预警。将知识图谱与监测预警平台、稳定性系数计算系统和趋势预测平台等 3 个关键信息系统对接,实现了多系统间的数据互通与功能协同,并显著提升了数据处理的实时性与分析的精确性。该系统根据未来 10 天的天气预报和库水调度数据,每日定时自动计算,输出三峡库区 4 518 个滑坡预警状态。

(5)实现了三峡库区监测预警的智能问答。所开发的系统基于三峡库区滑坡知识图谱,将滑坡相关的多源异构数据进行关联,构建了一个集知识整合、推理与解答于一体的智能问答机器人,它能回答库区滑坡数据查询、知识查询、统计分析、统计推理、预测预报多类信息,改变了以往需要调取多个信息系统才能获取信息的方式,本研究只通过人机对话即可完成。这种工作方式不但降低了用户的使用门槛,节省了用户的获取信息的时间,并且由于采用了检索增强生成技术,减少了大语言模型普遍存在的幻觉现象,所生成回答的质量较高。

虽然本研究将滑坡监测预警工作推向了知识化、智能化,但客观而言,本研究的工作也还比较初步,功能也比较原始,特别是大模型的建设方面还要进一步的应用和完善,下一步研究计划如下:

(1)继续完善库区地灾监测预警大模型。结合最新地质研究成果和最新发生的案例进行优化完善,利用实际监测数据和历史灾害数据对系统功能进行多场景、多角度验证,提高模型对复杂地质条件和新型灾害类型的适应性。

(2)进一步完善数据更新机制。收集、加工和投喂更多资料,及时纳入新发生地灾案例数据、更新调查成果等,确保数据时效性和准确性。

(3)进一步拓展应用场景。拓展知识图谱在地灾应急方案制定、灾后重建规划等方面的具体场景

应用,提升服务能力。

(4) 研究系统对声音、图像、图形等形式的输入和输出的支持。

(5) 对更多地灾信息系统进行智能化的改造,并开展基于 MCP 的智能体对接,让大模型从 chating 走向 acting。

附图见本刊官网 <https://doi.org/10.3799/dqkx.2025.127>。

References

- Abu-Salih, B., 2021. Domain-Specific Knowledge Graphs: A Survey. *Journal of Network and Computer Applications*, 185: 103076. <https://doi.org/10.1016/j.jnca.2021.103076>
- Cao, X. G., Zhang, M. Y., Lei, Z., et al., 2021. Construction and Application of Knowledge Graph for Coal Mine Equipment Maintenance. *Industry and Mine Automation*, 47(3): 41—45 (in Chinese with English abstract).
- Cheatham, M., Krisnadhi, A., Amini, R., et al., 2018. The GeoLink Knowledge Graph. *Big Earth Data*, 2(2): 131—143. <https://doi.org/10.1080/20964471.2018.1469291>
- Chen, L. J., Peng, L., & Yang, L. N., 2024. Improving Landslide Prediction: Innovative Modeling and Evaluation of Landslide Scenario with Knowledge Graph Embedding. *Remote Sensing*, 16(1): 145. <https://doi.org/10.3390/rs16010145>
- Du, Z. Q., Li, Y., Zhang, Y. T., et al., 2020. Knowledge Graph Construction Method on Natural Disaster Emergency. *Geomatics and Information Science of Wuhan University*, 45(9): 1344—1355 (in Chinese with English abstract).
- Fan, R. Y., Wang, L. Z., Yan, J. N., et al., 2020. Deep Learning-Based Named Entity Recognition and Knowledge Graph Construction for Geological Hazards. *ISPRS International Journal of Geo-Information*, 9(1): 15. <https://doi.org/10.3390/ijgi9010015>
- Feng, J., Xu, X., Lu, J. M., 2019. Construction and Application of Water Conservancy Information Knowledge Graph. *Computer and Modernization*, (9): 35—40 (in Chinese with English abstract).
- Gui, H., L. Yuan, L., Ye, H., et al., 2024. Iepile: Unearthing Large-Scale Schema-Based Information Extraction Corpus. <https://arxiv.org/pdf/2402.14710>
- Han, T., Huang, H. S., Yao, L. G., 2021. Entity Extraction for Aero-Engine Fault Knowledge Graph. *Modular Machine Tool & Automatic Manufacturing Technique*, (10): 69—73, 78 (in Chinese with English abstract).
- Hogan, A., Blomqvist, E., Cochez, M., et al., 2021. Knowledge Graphs. *ACM Computing Surveys*, 54(4): 1—37. <https://doi.org/10.1145/3447772>
- Li, X. P., Xu, J. H., Guo, Z. M., et al., 2019. Construction and Application of Knowledge Graph of Power Dispatching Automation System. *Electric Power*, 52(2): 70—77, 157 (in Chinese with English abstract).
- Liu, W. C., 2022. Construction and Application of Landslide Knowledge Graph (Dissertation). Hefei University of Technology, Hefei (in Chinese with English abstract).
- Lyu, H. K., Hong, L., Ma, F. C., 2020. Constructing Knowledge Graph for Financial Equities. *Data Analysis and Knowledge Discovery*, 4(5): 27—37 (in Chinese with English abstract).
- Ma, X. G., Ma, C., Wang, C. B., 2020. A New Structure for Representing and Tracking Version Information in a Deep Time Knowledge Graph. *Computers & Geosciences*, 145: 104620. <https://doi.org/10.1016/j.cageo.2020.104620>
- Pu, T. J., Tan, Y. P., Peng, G. Z., et al., 2021. Construction and Application of Knowledge Graph in the Electric Power Field. *Power System Technology*, 45(6): 2080—2091 (in Chinese with English abstract).
- Qi, G. L., Gao, H., Wu, T. X., 2017. The Research Advances of Knowledge Graph. *Technology Intelligence Engineering*, 3(1): 4—25 (in Chinese with English abstract).
- Qiu, Q. J., Wu, L., Ma, K., et al., 2023. A Knowledge Graph Construction Method for Geohazard Chain for Disaster Emergency Response. *Earth Science*, 48(5): 1875—1891 (in Chinese with English abstract).
- Raskin, R. G., Pan, M. J., 2005. Knowledge Representation in the Semantic Web for Earth and Environmental Terminology (SWEET). *Computers & Geosciences*, 31(9): 1119—1125. <https://doi.org/10.1016/j.cageo.2004.12.004>
- Ruan, T., Sun, C. L., Wang, H. F., et al., 2016. Construction of Traditional Chinese Medicine Knowledge Graph and Its Application. *Journal of Medical Informatics*, 37(4): 8—13 (in Chinese with English abstract).
- Singhal, 2012. Official Google Blog: Introducing the Knowledge Graph: Things, not Strings. <http://googleblog.blogspot.pt/2012/05/introducing-knowledge-graph-things-not.html>
- Wang, C. B., Ma, X. G., Chen, J. G., et al., 2018. Information Extraction and Knowledge Graph Construction from Geoscience Literature. *Computers & Geosciences*,

- 112: 112–120. <https://doi.org/10.1016/j.cageo.2017.12.007>
- Xu, Q., Cui, S. H., Huang, W., et al., 2023. Construction of a Landslide Knowledge Graph in the Field of Engineering Geology. *Geomatics and Information Science of Wuhan University*, 48(10): 1601–1615 (in Chinese with English abstract).
- Xu, Z. L., Sheng, Y. P., He, L. R., et al., 2016. Review on Knowledge Graph Techniques. *Journal of University of Electronic Science and Technology of China*, 45(4): 589–606 (in Chinese with English abstract).
- Zhang, X. Y., Huang, Y., Zhang, C. J., et al., 2022. Geoscience Knowledge Graph (GeoKG): Development, Construction and Challenges. *Transactions in GIS*, 26(6): 2480–2494. <https://doi.org/10.1111/tgis.12985>
- Zhang, Z. X., 2023. Stability Evaluation of High Slopes under Earthquake and Rainfall Based on Transfer Coefficient Method. *Shaanxi Water Resources*, (4): 14–15, 18 (in Chinese with English abstract).
- Zhou, C. H., Wang, H., Wang, C. S., et al., 2021. Research on Geo-Knowledge Map in the Age of Big Data. *Scientia Sinica (Terrae)*, 51(7): 1070–1079 (in Chinese).
- Zhou, Y. F., Duan, H., Zhao, H. L., et al., 2024. Hydrological Modeling Knowledge Graph Construction and Application. *Journal of Hydraulic Engineering*, 55(1): 80–91 (in Chinese with English abstract).
- Zhu, Y. Q., Sun, K., Wang, S., et al., 2023. An Adaptive Representation Model for Geoscience Knowledge Graphs Considering Complex Spatiotemporal Features and Relationships. *Science China Earth Sciences*, 66(11): 2563–2578. <https://doi.org/10.1007/s11430-022-1169-9>
- Zou, X. H., 2020. A Survey on Application of Knowledge Graph. *Journal of Physics: Conference Series*, 1487(1): 012016. <https://doi.org/10.1088/1742-6596/1487/1/012016>
- ### 中文参考文献
- 曹现刚, 张梦园, 雷卓, 等, 2021. 煤矿装备维护知识图谱构建及应用. *工矿自动化*, 47(3): 41–45.
- 杜志强, 李钰, 张叶廷, 等, 2020. 自然灾害应急知识图谱构建方法研究. *武汉大学学报(信息科学版)*, 45(9): 1344–1355.
- 冯钧, 徐新, 陆佳民, 2019. 水利信息知识图谱的构建与应用. *计算机与现代化*, (9): 35–40.
- 韩涛, 黄海松, 姚立国, 2021. 面向航空发动机故障知识图谱构建的实体抽取. *组合机床与自动化加工技术*, (10): 69–73, 78.
- 李新鹏, 徐建航, 郭子明, 等, 2019. 调度自动化系统知识图谱的构建与应用. *中国电力*, 52(2): 70–77, 157.
- 刘文聪, 2022. 滑坡知识图谱构建及应用(硕士学位论文). 合肥: 合肥工业大学.
- 吕华揆, 洪亮, 马费成, 2020. 金融股权知识图谱构建与应用. *数据分析与知识发现*, 4(5): 27–37.
- 蒲天骄, 谈元鹏, 彭国政, 等, 2021. 电力领域知识图谱的构建与应用. *电网技术*, 45(6): 2080–2091.
- 漆桂林, 高桓, 吴天星, 2017. 知识图谱研究进展. *情报工程*, 3(1): 4–25.
- 邱芹军, 吴亮, 马凯, 等, 2023. 面向灾害应急响应的地质灾害链知识图谱构建方法. *地球科学*, 48(5): 1875–1891.
- 阮彤, 孙程琳, 王昊奋, 等, 2016. 中医药知识图谱构建与应用. *医学信息学杂志*, 37(4): 8–13.
- 许强, 崔圣华, 黄维, 等, 2023. 面向工程地质领域的滑坡知识图谱构建方法研究. *武汉大学学报(信息科学版)*, 48(10): 1601–1615.
- 徐增林, 盛泳潘, 贺丽荣, 等, 2016. 知识图谱技术综述. *电子科技大学学报*, 45(4): 589–606.
- 张忠孝, 2023. 基于传递系数法地震及降雨作用下库岸高边坡稳定性评价. *陕西水利*, (4): 14–15, 18.
- 周成虎, 王华, 王成善, 等, 2021. 大数据时代的地质知识图谱研究. *中国科学: 地球科学*, 51(7): 1070–1079.
- 周逸凡, 段浩, 赵红莉, 等, 2024. 水文模型知识图谱构建与应用. *水利学报*, 55(1): 80–91.