

<https://doi.org/10.3799/dqkx.2025.268>



# 数据-知识协同驱动的共伴生矿产知识图谱构建方法

秦颖<sup>1</sup>, 杨慧<sup>1,2</sup>, 崔柳<sup>1</sup>, 酆格斐<sup>3</sup>, 王佳<sup>4</sup>, 乔亦娜<sup>1</sup>, 吕青宙<sup>1</sup>, 冯健<sup>5</sup>, 王文峰<sup>1,6</sup>

1. 中国矿业大学资源与地球科学学院, 煤层气资源与成藏过程教育部重点实验室, 江苏徐州 221116
2. 乌鲁木齐气象卫星地面站, 新疆乌鲁木齐 830002
3. 江苏师范大学语言科学与艺术学院, 江苏徐州 221009
4. 格林威治大学计算机与数学科学学院, 伦敦 SE109LS
5. 新疆维吾尔自治区地质局煤田地质中心, 新疆乌鲁木齐 830000
6. 新疆大学地质与矿业工程学院, 新疆乌鲁木齐 830047

**摘要:** 针对地质大数据与成矿知识割裂导致的共伴生关系建模难题, 亟需构建支撑智能分析的知识方法体系. 提出一种数据-知识协同驱动的知识图谱构建方法, 融合领域本体与 BERT-BiLSTM-CRF 模型, 通过“知识引导-数据反馈”机制实现本体演化与信息抽取的动态协同, 系统地从多源地质文本中提取矿床特征与共伴生关系, 建立勘查数据与成矿知识间的语义映射. 实验表明: 实体识别 F1 值达 83.2%, 较基线提升 15.4%; 实体重复率降低 5.7 个百分点, 图谱一致性显著改善. 最终构建包含 1.2 万节点与 2.8 万关系的结构化知识图谱, 支撑可视化分析、智能问答、成矿预测及平台服务. 该方法实现了知识与数据的深度融合, 为矿产勘查向数据-知识协同驱动的智能范式转型提供了可解释、可操作的技术路径.

**关键词:** 知识图谱; 本体; 数据知识协同; 共伴生矿产; 深度学习; Neo4j 图数据库.

中图分类号: P617; P628; TP18

文章编号: 1000-2383(2026)02-674-16

收稿日期: 2025-04-11

## Developing a Data-Knowledge Synergy-Driven Methodology for Co-Associated Minerals Knowledge Graph Construction

Qin Ying<sup>1</sup>, Yang Hui<sup>1,2</sup>, Cui Liu<sup>1</sup>, Feng Gefei<sup>3</sup>, Wang Jia<sup>4</sup>, Qiao Yina<sup>1</sup>, Lv Qingzhou<sup>1</sup>, Feng Jian<sup>5</sup>, Wang Wenfeng<sup>1,6</sup>

1. Key Laboratory of Coalbed Methane Resources & Reservoir Formation Process Ministry of Education, School of Resources and Geosciences, China University of Mining and Technology, Xuzhou 221116, China
2. Urumqi Meteorological Satellite Ground Station, Urumqi 830002, China
3. School of Linguistic Sciences and Arts, Jiangsu Normal University, Xuzhou 221009, China
4. School of Computing and Mathematical Sciences, University of Greenwich, London SE109LS, United Kingdom
5. Coalfield Geological Center, Geological Bureau of Xinjiang Uygur Autonomous Region, Urumqi 830000, China
6. School of Geology and Mining Engineering, Xinjiang University, Urumqi 830047, China

**基金项目:** 国家自然科学基金面上项目(Nos. 42571545, 52478011); 第三次新疆综合科学考察项目(No. 2022xjkk1006); 新疆维吾尔自治区重点研发项目(No. 2022B01012-1); 江苏省自然资源科技计划项目(No. 2023018); 中央高校基本科研业务费专项资金资助(No. 2024ZDPYCH1002); 江苏省科技智库计划项目(No. JSKX0225042).

**作者简介:** 秦颖(2002-), 女, 硕士研究生, 地球探测与信息技术专业, 主要从事矿产资源大数据智能挖掘研究. ORCID: 0009-0008-1683-943X. E-mail: qinying@cumt.edu.cn

**\*通讯作者:** 杨慧, ORCID: 0000-0001-9421-3573. E-mail: yanghui@cumt.edu.cn

**引用格式:** 秦颖, 杨慧, 崔柳, 酆格斐, 王佳, 乔亦娜, 吕青宙, 冯健, 王文峰, 2026. 数据-知识协同驱动的共伴生矿产知识图谱构建方法. 地球科学, 51(2): 674-689.

**Citation:** Qin Ying, Yang Hui, Cui Liu, Feng Gefei, Wang Jia, Qiao Yina, Lv Qingzhou, Feng Jian, Wang Wenfeng, 2026. Developing a Data-Knowledge Synergy-Driven Methodology for Co-Associated Minerals Knowledge Graph Construction. *Earth Science*, 51(2): 674-689.

**Abstract:** The growing disconnect between geological big data and metallogenic knowledge poses a significant challenge to modeling co-associated mineral relationships, underscoring the urgent need for a knowledge-based methodology capable of supporting intelligent analysis. To address this, we propose a data-knowledge synergy-driven approach for constructing knowledge graphs, which integrates domain ontology with the BERT-BiLSTM-CRF model. By leveraging a “knowledge-guided, data-informed” mechanism, the method enables dynamic collaboration between ontology evolution and information extraction, systematically identifying ore deposit features and co-associated relationships from multi-source geological texts and establishing semantic mappings between exploration data and metallogenic knowledge. Experimental results show that entity recognition achieves an F1 score of 83.2%, representing a 15.4 percentage-point improvement over the baseline; entity redundancy is reduced by 5.7 percentage points, markedly enhancing graph consistency. The resulting structured knowledge graph, which comprises 12,000 nodes and 28,000 relations, has been deployed in visualization analysis, intelligent question answering, mineralization prediction, and data platform services. This work realizes deep integration of data and knowledge, offering an interpretable and actionable technical pathway for transforming mineral exploration from an experience-driven paradigm to one driven by data-knowledge synergy.

**Key words:** knowledge graph;ontology;data-knowledge synergy; symbiotic and associated minerals; deep learning; Neo4j graph database.

## 0 引言

矿产资源的高效勘探与开发是国家资源安全与经济可持续发展的重要保障。资源勘查过程中积累的海量矿产地质数据具有显著多源异构性和语义歧义性(赵鹏大,2019;左仁广,2021),导致原始矿产地质信息向可推理的专业成矿知识的转化困难,制约了数据在矿产预测等智能分析流程中的有效应用(周永章和肖凡,2024)。共伴生矿产作为主矿产开采过程中的关键性伴生有价值组分(吴占伟等,2024),其矿物组合间存在着典型的空间关联、区域差异与动态演化特征,为深入解析成矿规律和提升资源综合利用效能提供了重要依据(鞠建华等,2024)。然而,由于这类复杂知识的系统化建模需求与多源异构数据整合分析的瓶颈制约,亟需构建系统化的矿产资源知识管理体系,以驱动资源评价与找矿预测模式向智能化方向发展(周成虎等,2021;王成彬等,2024;吴润泽等,2025;张宝一等,2025)。

知识图谱凭借其独特的结构化语义网络表达能力,通过对概念、实体及其语义关联进行符号化建模,构建富含属性的实体关系知识网络(田玲等,2021;韩伟等,2025)。其在数据集成、语义解析与智能决策支持方面展现出显著优势,为矿产资源知识管理提供了新路径(Yu *et al.*, 2025)。近年来,相关研究已取得初步进展:例如,基于网络爬虫整合多源数据,构建涵盖地质实体、地理区域与组织机构的知识图谱(王晴等,2022);依托成矿系统概念模型实现矿床知识特征的系统化表达(周永章等,

2021);融合地质时空本体以增强矿产预测的语义建模能力(叶育鑫等,2024);引入BERT等深度学习模型提升矿床实体识别性能(汪陈,2021);以及结合自然语言处理与机器学习技术,推动矿产勘查知识图谱的自动化构建与语义推理(张前龙等,2024)。

然而,当前矿产资源领域知识图谱仍普遍存在明显局限:一方面,建模方法往往侧重于单一驱动方式,数据驱动方式因缺乏领域知识约束,易产生语义偏差;知识驱动方式则受限于人工构建的高成本与滞后性。另一方面,现有研究多聚焦于单一矿种或典型矿床类型,对复杂共伴生矿产体系的系统性建模仍显薄弱,尚未形成统一的建模框架,难以有效刻画其独特的空间依存、组合共生与演化规律。因此,亟需有机融合数据驱动与知识驱动的优势,构建面向共伴生矿产的结构化、可推理知识图谱,实现更精准的矿产知识表达与智能应用。

为此,本文聚焦共伴生矿产知识体系,提出一种数据-知识协同驱动的知识图谱构建范式:融合领域本体建模与基于BERT-BiLSTM-CRF的混合信息抽取技术,并引入语义对齐与人工校验协同的知识融合机制,实现多源异构地质数据向结构化成矿知识的高效转化,建立勘查大数据与成矿知识体系的语义映射,构建共伴生矿产知识图谱。通过“知识引导-数据反馈”的闭环协同,智能提取地质构造特征、成矿规律及矿物伴生关系等核心知识,为矿产资源识别、矿体定位预测与勘查精度提升提供可解释、可复用的语义支撑。

表 1 矿产领域数据类型

Table 1 Data types in mineral resource domain

数据类型	典型形式	语义特征
结构化数据	统计表单、关系型数据库等	预定义字段与关系,严格模式约束
半结构化数据	XML/JSON 报告、标签化地质图件等	嵌套层级,部分字段标记
非结构化数据	地质调查报告、科研论文、勘查日志等	自由文本,隐含领域知识

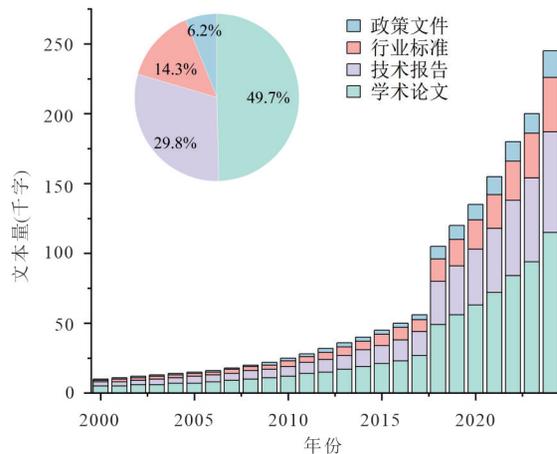


图 1 原始语料文本类型分布

Fig.1 Distribution of text types in the original corpus

## 1 数据预处理与语料库构建

### 1.1 数据来源与预处理

矿产信息资源按领域知识表达形式可分为 3 类:结构化数据(如统计表单、数据库)具有固定模式和明确字段,便于处理但表达灵活性差;半结构化数据(如标签图件)通过层级和标签组织信息,兼具一定结构与扩展性;非结构化数据(如论文、报告、日志)以自由文本承载丰富语义,知识密度高但缺乏统一格式,难以自动解析.表 1 简要概括了 3 类数据的典型形式与核心语义特征,突显其在结构化程度、语义表达与处理难度上的差异.

由于部分核心地质勘查数据的获取权限限制,本研究以科技文献及网络公开资源中的非结构化文本作为主要数据源.通过智能信息采集技术与文献筛选相结合,系统采集了 2000—2024 年间发布的政策文件、行业标准、技术报告 and 学术论文等非结构化文本,构建的原始语料规模达 150 万字符.如图 1,语料构成分析显示学术论文(49.7%)和技术报告(29.8%)占比最高,合计近 80%,构成研究主体;行业标准和政策文件分别占 14.3% 和 6.2%.时序分析表明,2000—2024 年间 4 类文档的文本量均呈现稳定增长态势,其中学术论文年均增长率达 14.0%,

技术报告为 13.8%,充分反映出矿产领域技术信息持续加速积累的趋势,为本研究的深度知识挖掘提供了动态、丰富且持续更新的数据支撑.

在数据预处理阶段,对原始语料进行了系统性清洗与规范化处理,包括文本格式统一、冗余信息剔除以及非地质领域内容和敏感信息的过滤.经过标准化预处理流程,构建了约 115 万字的高质量文本数据集.该数据集经过校验与整理,具备规范的结构化特征和良好的可扩展性,能够有效支持后续的语料库构建、命名实体识别及关系抽取等自然语言处理任务.

### 1.2 语料库构建

尽管本体构建尚缺乏普适性的方法论,但其作为领域知识的形式化规范与共享基础(杨慧等, 2008),在专家共识度较高的地质矿产领域,因知识体系成熟、术语规范性强,更易于构建高质量本体.为支撑共伴生矿产知识结构化组织与可计算化的语义表达,本文提出融合数据反馈机制的协同式本体建模方法,该方法采用自顶向下(top-down)的建模策略,通过系统整合成矿理论、矿产资源行业标准及领域与专家认知,构建层级清晰的本体知识框架体系,并引入勘查报告、剖面图、矿物组合记录等多源地质调查数据进行验证与反馈,形成“知识引导建模—数据反馈优化”的协同建模策略,作为“知识引导—数据反馈”机制在本体构建阶段的具体实现路径,如图 2 所示.

本方法流程包含 3 个相互关联的阶段.首先,基于成矿理论与行业规范开展本体建模,通过系统梳理共伴生矿产相关的核心概念及其属性关系,构建初步的本体结构;其次,设计面向多维语义表达的标注框架,并结合 Protégé 与 LabelStudio 等工具完成语料标注;最后,标注过程中同步引入冲突检测与一致性检验机制,用于识别本体中存在的概念缺失或语义模糊问题.这些问题被反馈至初始建模环节,形成从标注实践到本体优化的闭环,推动模型的动态修订与持续完善.该协同机制实现了领域知识体系与实际语料之间的双向互动,在提升本体理

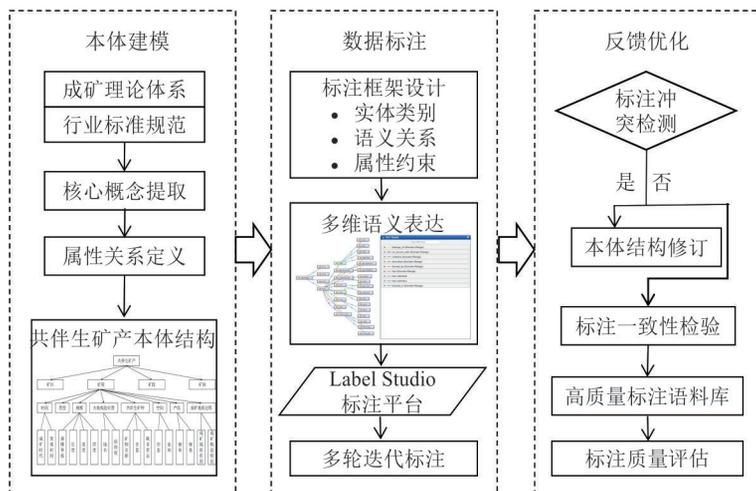


图 2 知识引导与数据反馈的协同建模流程

Fig.2 Flowchart of the knowledge-driven and data-enhanced collaborative modeling processes

表 2 共伴生矿产本体五元组结构

Table 2 Five-tuple structure of the co-associated minerals ontology

所属本体	成分	符号表示	含义说明
CM <sub>Ontology</sub>	概念 (Concept)	CM <sub>Concept</sub>	领域核心术语集合,如矿区、矿床、矿段、矿体等
	属性 (Property)	CM <sub>Property</sub>	实体特征信息,如成矿年代、构造位置、矿体厚度等
	关系 (Relation)	CM <sub>Relation</sub>	实体间语义关联,如共生、包含、继承、实例等
	规则 (Rule)	CM <sub>Rule</sub>	约束矿种组合与成因联系的逻辑规则
	实例 (Instance)	CM <sub>Instance</sub>	概念到具体对象的映射,如“金堆城钨矿”对应“斑岩型钨矿”类型

论完备性的同时,也增强了其在具体应用场景中的适应能力.

最终构建的共伴生矿产语料库结构规范、语义清晰,为后续知识抽取与知识图谱构建奠定了高质量的数据基础.

**1.2.1 知识与本体建模** 参照斯坦福七步法构建领域本体 (Du *et al.*, 2024),其标准化流程包括:①确定本体领域和范围;②复用现有本体资源;③枚举本体核心术语;④定义类与类层级体系;⑤定义类的属性(槽位);⑥定义属性约束(槽位刻画);⑦创建类实例.为顺应语义 Web 发展需求,本研究将上述方法论输出抽象为“概念—属性—关系—规则—实例”五元组结构,并基于 OWL 语言实现形式化建模,其核心要素如表 2 所示.该本体具备良好的语义表达能力与机器可读性,支持后续的知识推理、查询扩展与跨系统共享,有效推动地质知识向语义 Web 环境的融合.

在具体实现中,OWL 通过类 (Class) 对应五元组中的“概念”,用于定义矿区、矿体等核心地质实

体;通过数据属性 (DataProperty) 和对象属性 (ObjectProperty) 分别表达“属性”与“关系”,如矿体厚度、共生关系等,实现特征描述与语义关联的结构化建模.逻辑规则部分则利用 OWL 的公理机制(如类包含、属性限制等)对成矿组合、空间依存等关系形式化约束;对于涉及多条件组合的复杂规则,则结合 SWRL 规则语言(支持“若...则...”形式的逻辑推理)进行补充建模 (Zhang and Lobov, 2024; Bilenchi *et al.*, 2025).同时,将实际勘查数据中的具体对象作为个体 (Individual) 实例化到相应概念下,实现概念到实例的具体落地.

**1.2.2 数据标注与质量评估** 为提升通用语言模型在共伴生矿产领域知识抽取中的准确性,本文基于所构建的领域本体,依托 Protégé 建模环境,设计了一套语义结构清晰、层级明确的标注体系.该体系以本体中的类、数据属性和对象属性为语义锚点,通过建立文本片段与本体元素之间的语义映射关系,系统覆盖矿产实体在成矿地质背景、空间展布特征、成矿演化过程、规模品位信息及共伴生组

表 3 共生矿产实体多维语义特征分类

Table 3 Multi-dimensional semantic feature classification of co-associated mineral entities

维度	子类目	描述说明
时间	成矿时代	指矿产形成所处的具体地质历史阶段,通常以地质纪或世等单位表示,如侏罗纪晚期、早白垩世等
	发现时间	矿产地首次在文献中被记录的时间
大地构造位置	地名	矿产所在的具体地理区域或矿区名称,用于标识其空间归属
	经纬度	标注矿产实体中心位置的地理经纬度信息,用于精确定位
类型		根据成因、矿物组合及构造背景对矿床进行分类
成矿地质过程	成矿地质作用	矿床形成的地质作用机制,包括内生、外生和变质成矿作用等
	成矿构造特征	控制矿体形成、定位和分布的主要构造要素,如断裂带、褶皱构造等
空间	走向	描述矿体在空间上的延伸方向
	形态	描述矿体在空间上的展布形态,如似层状、脉状、透镜状等
产状	倾向	矿产实体倾斜面的水平投影方向,用于表征其空间产出方位
	倾角	矿产实体倾斜面与水平面之间的夹角,反映其倾斜程度
规模	规模等级	对矿体资源量和体量的定性划分,分为大型、中型、小型等
	长度	矿体沿某一方向延伸的长度,用于定量描述其几何规模
	宽度	矿体横向扩展的宽度,是衡量矿体规模的重要参数之一
	厚度	矿体在垂直或倾斜方向上的厚度,影响矿产开采价值和经济性
共生矿产	矿物名称	矿产实体中主要共存或伴生的金属或非金属元素种类
	含量	元素在矿石中的平均含量,常以百分比(%)或克/吨(g/t)表示
	共生元素赋存状态	元素在矿物中的存在形式,如独立矿物、类质同象、吸附态、包裹体等形式

合关系等维度的多粒度表达.该映射机制有效支撑了非结构化文本向结构化、机器可读知识的转换,为后续知识抽取提供了可解释的语义框架,具体内容如表 3 所示.

在标注体系实施过程中,采用“知识引导—数据反馈”的协同机制,并实施“基于知识共识的标签定义—面向真实语境的迭代优化”策略.一方面,基于成矿理论、行业标准及领域专家对勘查术语的认知共识,系统定义实体类别与语义关系,体现知识对标注体系的引导作用;另一方面,引入多源地质调查数据作为验证语料,通过实际文本中的标注实践发现问题、修正歧义,实现对标签体系的动态优化,体现数据对模型构建的反馈价值.语料标注工作依托 Label Studio 平台,采用多轮迭代与自我校验相结合的方式,在持续的冲突检测与标准统合中不断完善标签体系与标注规范,形成从理论指导到实践反哺的闭环演化过程.

最终选取具有代表性的 10% 语料(约 4 750 条句子)完成精细化标注,涵盖综述类文献、研究报告及学位论文等多种文本来源.为全面评估标注体系的质量与可操作性,从一致性、覆盖率、标签迭代情况及使用频率四个方面开展多轮标注实验.结果表明,该体系具有良好的一致性、覆盖性和实用性,能够有效支持后续信息抽取与知识图谱构建的需要.

标注一致性方面,采用 Cohen's Kappa 系数(Pérez *et al.*, 2020)进行度量,其数学表达式为:

$$K_{\text{Cohen}} = \frac{P_o - P_e}{1 - P_e}, \quad (1)$$

其中: $P_o$ 表示两次独立标注的观察一致率; $P_e$ 为基于类别分布计算的随机一致概率.该指标有效排除了偶然一致的影响,更准确反映标注者间的实际一致性.经 3 轮迭代标注,并对 5% 样本进行复标验证,最终  $K=0.85$ ,表明标注协议具有高度稳定性和良好可操作性.

实体类型覆盖率的计算如式(2):

$$R_{\text{coverage}} = \frac{N_{\text{labeled}}}{N_{\text{total}}} \times 100\%, \quad (2)$$

其中: $N_{\text{labeled}}$ 为已标注的实体类型数量; $N_{\text{total}}$ 为本体中定义的实体类型总数.结果显示,14 类本体实体中已有 13 类被实际标注,覆盖率达 93%,验证了标注体系对领域语义空间的充分表达能力.

在迭代优化过程中,经过三轮标注共新增标签 4 项、修订 6 项,体现了标注体系良好的动态适应与持续完善能力.

标签使用频率计算如式(3):

$$f_i = \frac{C_i}{C_{\text{total}}} \times 100\%, \quad (3)$$

其中: $f_i$ 表示第  $i$  类标签的使用频率, $C_i$ 为该类实体

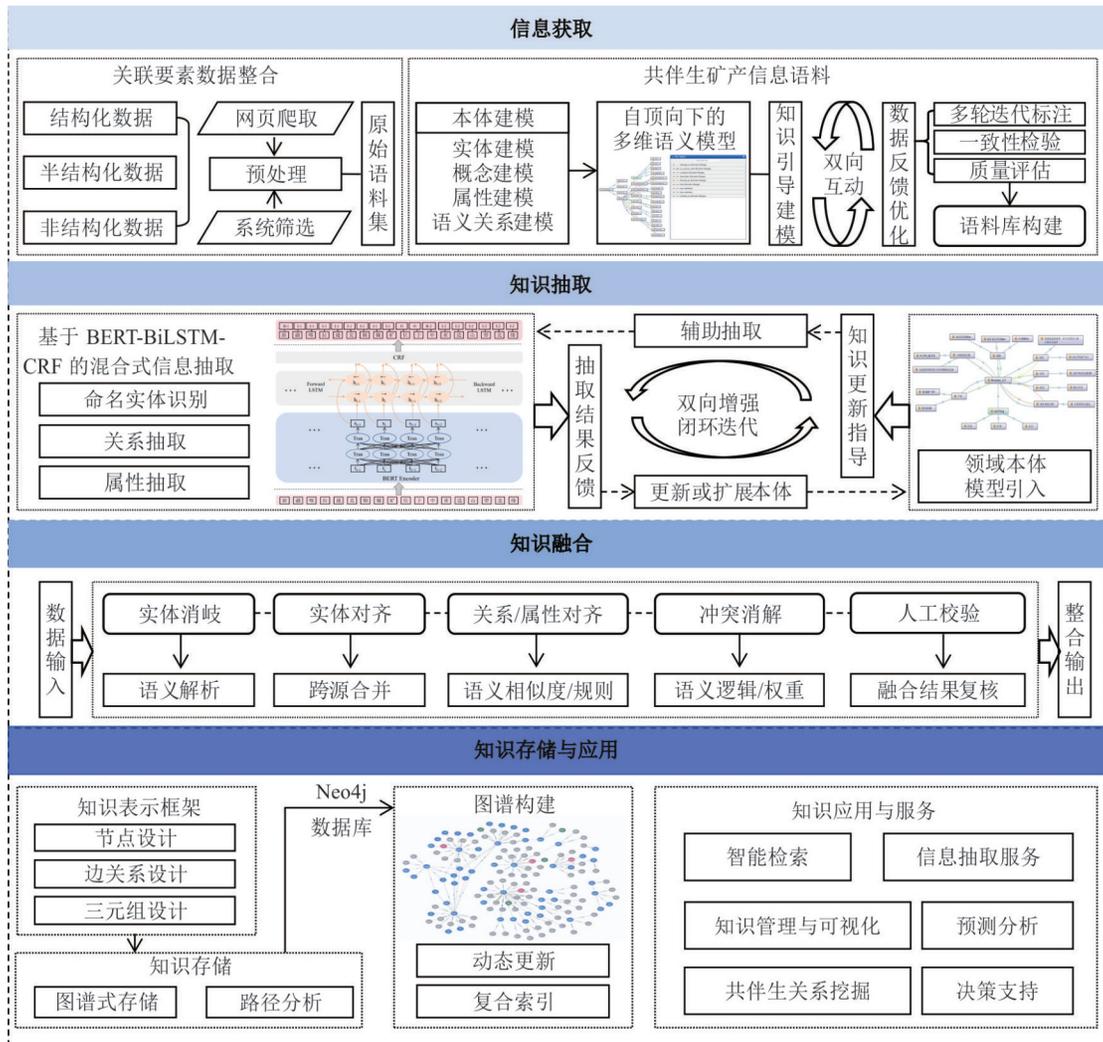


图 3 共生伴生矿产知识图谱技术路线图

Fig.3 Technical roadmap for the construction of co-associated mineral resources knowledge graph

在语料中出现的次数,  $C_{total}$  为所有实体标注的总次数. 分布分析表明,大地构造位置(34%)和共生伴生矿种(31%)为最高频类别. 该分布特征契合地质文本中对控矿要素与矿化组合的重点描述需求.

## 2 研究方法

面向共生伴生矿产资源智能化勘查的实际需求, 本文提出一种数据—知识协同驱动的知识图谱构建方法, 通过深度融合多源异构地质数据与成矿领域知识, 实现数据驱动表征与知识引导推理的有机协同, 旨在将非结构化文本转化为结构化、可推理的知识体系. 该方法以“知识引导—数据反馈”为核心协同机制, 在知识建模与信息处理的全过程中实现领域知识与实际数据的双向互动: 知识体系指导信息抽取与语义建模, 实际数据则通过标注与融合

过程反向优化知识模型, 推动其动态演进.

整体技术路线涵盖信息获取、知识抽取、知识融合与知识存储及应用4个环节(图3). ①信息获取阶段通过构建领域语料库完成多源文本的系统采集; ②知识抽取阶段采用知识增强型BERT-BiLSTM-CRF与BERT关系抽取框架, 融合上下文语义表征与领域规则, 实现地质实体及其实体间关系与属性关联的精准识别; ③知识融合阶段引入实体对齐与规则驱动的后处理机制, 结合向量语义匹配与人工校验, 解决多源表述异构性带来的实体冗余与语义冲突问题; ④知识存储与应用层面, 采用W3C标准的RDF模型与OWL语言对矿产概念、属性约束及逻辑规则进行形式化建模, 保障知识的语义一致性与可计算性; 最终基于Neo4j图数据库完成知识的结构化存储与高效查询, 兼具复杂关系检索与系统扩展能力, 服务于矿产知识的智能管理与

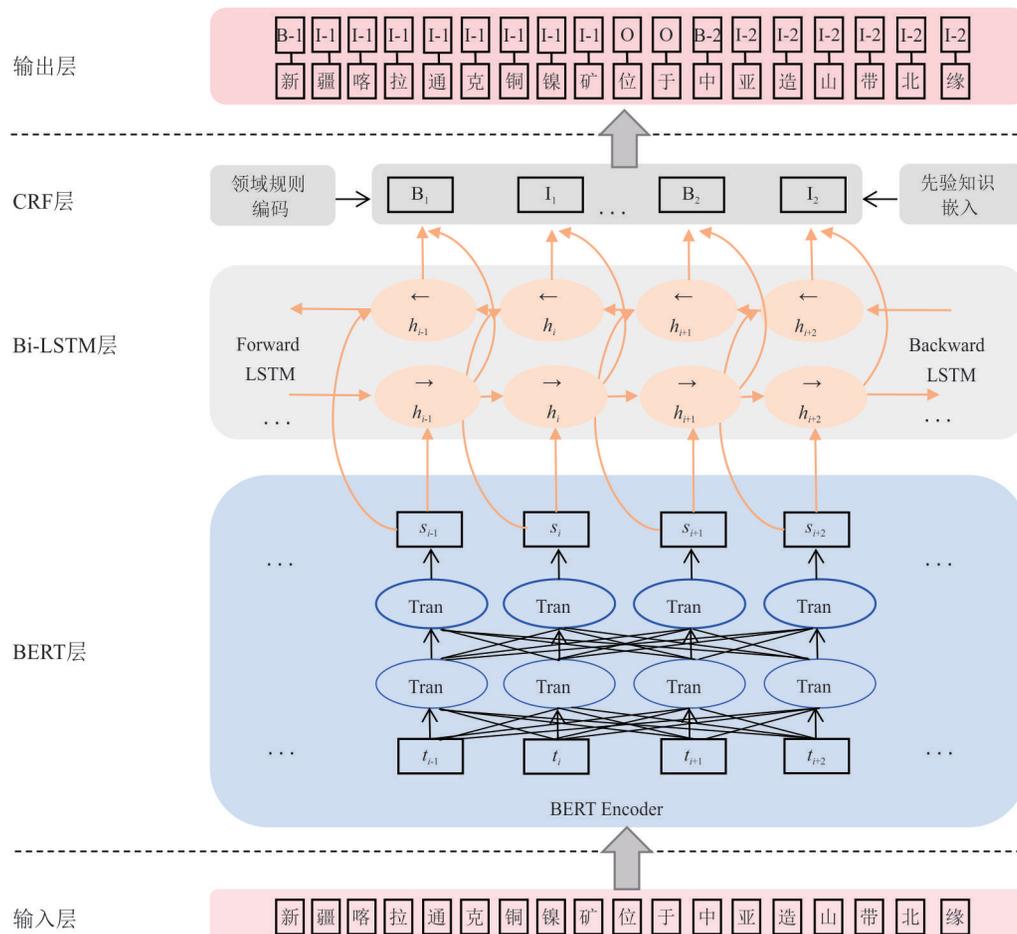


图4 BERT-BiLSTM-CRF混合抽取模型

Fig.4 BERT-BiLSTM-CRF hybrid extraction model

辅助决策。

## 2.1 知识抽取

在共伴生矿产知识体系深化过程中,由于非结构化文本具有语义跨度长、句法结构复杂和术语密度高等特点,精准识别实体、属性及关系的高质量知识抽取成为构建结构化知识的关键环节(廖振鑫, 2021)。传统知识抽取中基于规则与词典的方法依赖人工构建的语义资源,虽具有一定的可解释性和可控性,但由于泛化能力较弱,难以适应非结构化文本中多样化的表达形式(李冬梅等, 2020)。半监督学习策略可在标注数据有限的情况下提升模型训练效果,但其性能高度依赖于初始种子集的质量,因而对研究者的专业判断提出了较高要求(Bull *et al.*, 2018)。近年来,深度学习技术的快速发展使得基于神经网络的语言模型在自然语言处理领域取得显著突破,为共伴生矿产信息的自动化抽取提供了新的技术解决方案,该类方法能有效捕捉文本上下文语义特征,并在命名实体识别、属性抽取及

关系分类等任务中表现出优于传统方法的性能(鄂海红等, 2019)。

**2.1.1 实体抽取** 现有研究多采用单一神经网络模型进行地质实体与属性的信息提取,难以全面捕捉共伴生矿产文本中复杂的语义特征(Qiu *et al.*, 2019; 徐浩然等, 2025)。针对共伴生矿产数据多样性与领域知识异构性导致的语义理解局限(赵汀等, 2025),本文提出一种数据—知识协同驱动的多层次知识抽取框架。该框架并非传统意义上的单向流水线,而是构建了一个“抽取反馈—本体更新—知识反哺”的动态闭环,作为“知识引导—数据反馈”协同机制在信息抽取环节的关键执行单元,确保抽取过程兼具领域适应性与知识体系的动态扩展能力。

具体实现上,实体抽取采用知识增强型BERT-BiLSTM-CRF架构(图4),深度融合深度学习技术与领域规则知识:在数据驱动端,采用动态融合领域词典的BERT模型生成上下文感知表征,结合

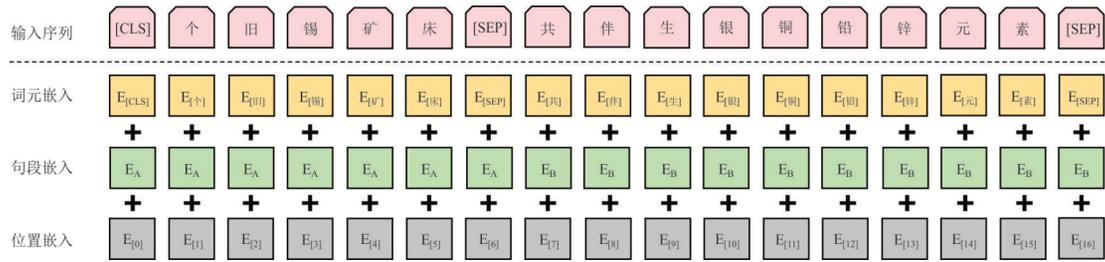


图 5 基于BERT的语义关系抽取模型

Fig.5 BERT-Based semantic relation extraction model

BiLSTM-CNN挖掘序列模式特征;在知识引导端,设计基于CRF的规则编码器,将矿产命名构词规则、先验知识以及手工定义的标签转移约束结构化嵌入解码过程。

① BERT (bidirectional encoder representations from transformers)层作为模型的底层语义编码模块,负责将输入文本映射为具有丰富语义信息的上下文感知向量表示(谢腾等,2020)。该层基于预训练语言模型对输入文本进行逐词编码,能够有效捕捉词与词之间的长距离依赖关系及语义上下文信息,为后续序列建模提供高质量的特征表示(王乃钰等,2021),尤其适用于处理共伴生矿产文本中的复杂地质术语、多义词及专业表达。

② Bi-LSTM (bidirectional long short-term memory)层用于对BERT输出的上下文向量进行序列建模,进一步提取文本中的时序语义特征(Yin *et al.*, 2025)。通过前向与后向LSTM网络的联合建模,能够同时捕捉当前词在前后文中的语义依赖关系,增强模型对复杂句法结构和嵌套语义表达的理解能力,从而在实体识别任务中有效识别实体边界、判断实体类型,并为后续标签解码提供上下文一致的特征表示。

③ CRF (conditional random field)层作为模型的顶层解码模块,负责对BiLSTM输出的特征向量进行标签预测,并通过引入标签之间的转移约束实现标签序列的全局优化(陈伟等,2018)。该层不仅能够建模标签间的依赖关系(如“B-实体”后不能直接接“I-其他实体”),还通过嵌入领域专家制定的先验规则(如基于领域知识设计的转移矩阵),实现对矿产领域术语结构的增强识别。

**2.1.2 关系抽取** 针对共伴生矿产文本中复杂的语义关联,构建了一种基于BERT的关系抽取框架(图5),旨在识别实体间的核心关系(如“共生”“伴生”)以及实体与其属性之间的语义联系(Li *et al.*,

2022)。该方法依托BERT强大的上下文建模能力(Meng *et al.*, 2022),能够动态融合前后语义信息,对地学文本中的复杂句法结构和专业术语进行深度语义编码(庄子浩等,2024),有效捕捉诸如构造关系、矿物共生关系及成矿环境间的因果联系等关键语义模式。为进一步增强模型对领域特有关系的理解能力,本模块引入领域专家定义的规则与先验知识,在模型微调阶段融入知识引导,优化关系分类边界。

## 2.2 知识融合

本文提出的混合信息抽取框架在共伴生矿产语义关系识别中初见成效,但仍存在实体边界模糊与关系类型混淆等问题,需进一步优化。例如,“铁矿—铜矿—共生”等复合型实体关系的三元组边界难以精准界定,且“伴生”与“共生”关系易因表达歧义而误判。经误差溯源分析,问题根源可归结为3方面:地质文本表达多样性导致同一关系存在多变表述(如“共生”被表述为“共同产出”或“相伴而生”);部分关系具有隐含性,需依赖地质背景知识(如年代、岩层特征)进行推理;标注数据中稀有关系(如伴生关系仅占训练集12.3%)样本量不足,限制了模型对特定关系的判别能力。因此,单一信息抽取模块难以保障知识图谱构建的一致性与完整性。

为应对上述问题,本文引入知识融合机制作为后处理优化的关键环节。在地学知识图谱构建过程中,多源文献的表述异构性(如同一矿床被表述为“个旧锡矿床”与“个旧锡铜多金属矿床”)不仅加剧了实体识别的难度,也进一步放大了信息抽取结果中的边界模糊与关系误判问题。为此,本文一方面通过动态扩充与优化标注语料库增强模型泛化能力;另一方面,依托知识融合技术中的实体对齐与人工校验机制,对抽取结果进行后处理以提升准确性。在数据清洗阶段,综合采用同义词词典、语义匹配算法及基于规则的实体匹配策略,对重复或变体

实体进行归并与表述标准化,从而显著提升数据存储效率与知识图谱的可操作性。

知识融合的核心在于对多源异构实体进行语义归一化。本文通过实体消歧、实体对齐、属性/关系对齐及冲突消解等关键技术,统一结构各异但语义一致的实体。具体而言,以向量化表示为基础,结合语义相似度计算(见式 4、式 5),将共伴生矿产实体映射至统一特征空间。当实体特征向量的余弦相似度超过经验阈值 0.9 时,判定为同一实体并执行合并操作,同时引入人工校验机制,确保对齐结果的准确性与知识融合的可靠性(杨秀璋,2016)。

$$V = \alpha \cdot V_{\text{char}} + \beta \cdot V_{\text{word}} + \gamma \cdot V_{\text{ctx}}, \quad (4)$$

$$V_{\text{ctx}} = \frac{1}{N(c)} \sum_{f \in c} W_f, \quad (5)$$

式中:  $V_{\text{char}}$  表示为基于实体名称字符序列生成的字符级向量;  $V_{\text{word}}$  为由领域微调语言模型输出的词级向量;  $V_{\text{ctx}}$  表示实体上下文语境的平均语义向量特征向量;  $W_f$  是上下文词  $f$  对应的词向量;  $N(c)$  为上下文窗口中的词项总数;  $\alpha, \beta, \gamma$  为归一化融合权重,满足  $\alpha + \beta + \gamma = 1$ 。

### 2.3 知识存储与应用

图数据库以节点和边的形式直观表达实体间的复杂关联,是构建知识图谱的理想载体。当前主流图数据库包括 Neo4j、JanusGraph、HugeGraph 和 Dgraph 等(杨军,2020;邱宇辰,2021)。相较于传统关系型数据库,图数据库在处理高连接度数据方面具有显著优势,尤其适用于需频繁执行多跳查询与复杂推理的知识图谱场景。针对共伴生矿产知识图谱中实体高度互联的特点,本文选用 Neo4j 作为底层存储引擎。该平台采用原生图存储机制,支持高效的图遍历操作,并允许以灵活方式建模图结构,在组织、检索和扩展复杂地质关系方面具有明显优势。同时,其内置的事务管理机制与对属性图模型的良好支持,有助于保障知识图谱运行的稳定性与后续扩展的可行性(王红等,2017)。

在知识表示层面,本文构建了面向矿产领域的结构化表示框架。该框架以三元组(subject-predicate-object)为基础语义单元,并将其映射为图结构中的节点(node)、边(relationship)和属性(property)。具体而言,节点对应矿床、矿体等地质实体;边用于表达“共生”“伴生”“赋存于”等地质关系;属性则用于描述实体或关系的附加信息,如品位、储量、空间位置等。这种表示模式不仅契合图数据库的建

模逻辑,也为后续的知识推理与查询分析提供了统一的数据接口。

在数据存储实现中,本文将融合后的三元组数据批量导入 Neo4j 数据库,并采用基于唯一标识符的去重机制确保实体一致性。具体利用 Python 脚本调用 Neo4j 官方驱动,结合 Cypher 语言中的 MERGE 操作与唯一约束策略,实现判断性插入,有效避免重复节点的生成。同时,在如矿床名称、矿物种类等关键字段上建立索引,并引入复合索引机制以优化多条件组合查询的执行效率。

为提升图谱的可用性与扩展性,本文还设计了动态更新策略。通过周期性执行预定义的 Cypher 脚本,支持知识图谱的增量式扩展,确保矿产知识库的持续更新与时效性,满足地质数据动态变化的研究需求。

## 3 结果与分析

### 3.1 知识抽取结果

为验证模型性能,本文对比了 3 种主流架构:传统的 BiLSTM-CRF 基线模型、基于预训练语言模型的 BERT-CRF 架构、以及本文采用的 BERT-BiLSTM-CRF 混合模型。所有模型均在同一标注数据集上进行训练与评估。数据按 7:1:2 的比例随机划分为训练集(3 325 条)、验证集(475 条)和测试集(950 条)。实验采用信息抽取领域广泛使用的 3 项核心评价指标,即准确率(precision,  $P$ )、召回率(recall,  $R$ )以及 F1 值(F1-score),以系统评估领域中实体与属性信息抽取的性能。准确率用于衡量模型预测结果中正确识别实体或属性占预测结果的比例,反映模型的精确性;召回率则表示在实际存在的实体或属性中,被成功识别出来的比例,体现模型抽取的完整性;F1 值通过对准确率与召回率进行调和平均计算,综合反映模型的整体性能。

实验结果(表 4)表明,本文提出的多特征融合模型在实体识别和属性识别任务上均显著优于对比方法。在实体识别任务中,本模型 F1 值达到 83.2%,较基线 BiLSTM-CRF 提升 15.4 个百分点;属性识别任务 F1 值为 80.3%,提升 11.9 个百分点。特别值得注意的是,模型在实体识别的召回率表现尤为突出( $R=86.7\%$ ),表明其能更全面地捕捉矿产文本中的各类实体。

为评估关系抽取模块在共伴生矿产语义识别中的有效性,本文从测试集中随机选取样本进行结

表4 实体与属性信息抽取结果

Table 4 Extraction results of entities and attributes information

模型	任务指标					
	实体识别			属性识别		
	precision	recall	F1-score	precision	recall	F1-score
BiLSTM-CRF	67.9%	67.7%	67.8%	60.6%	57.9%	59.2%
BERT-CRF	76.9%	73.3%	75.1%	68.1%	68.8%	68.4%
BERT-BiLSTM-CRF	79.9%	86.7%	83.2%	80.1%	80.6%	80.3%

表5 关系抽取结果示例

Table 5 Relation extraction results

原始文本	实体1	实体2	预测关系	真实标签	是否正确
金堆城钼矿床主要共伴生矿种有黄铁矿、黄铜矿	黄铁矿	黄铜矿	共生	共生	是
个旧锡矿位于云南东南有色金属成矿带西端,矿床中锡、铜等元素呈现共伴生形态	锡	铜	共生	共生	是
当前已经完成了矿区的详查工作,在钽铌铷铍矿中发现矿石中还含有大量的共(伴)生矿产,如铍和铷	铍	铷	伴生	伴生	是

表6 知识融合前后关键指标对比

Table 6 Comparison of key metrics before and after knowledge fusion

指标	知识融合前	知识融合后	变化幅度
实体重复率(重复实体数量占比)	11.40%	5.70%	↓5.7%
实体表述标准化率	67.20%	83.50%	↑16.3%
知识图谱节点数量(实体总数)	12 553	10 800	↓14.0%
知识图谱边数量(关系总数)	34 200	32 100	↓6.1%
人工校验工作量(需审核实体对数量)	2 522	1 104	↓56.2%

果分析.表5列出了模型在若干典型文本上的关系抽取结果示例,反映了其在复杂句式和多实体场景下的输出情况.实验结果表明,模型在完整测试集上表现出较高的整体性能,精确率达到85.6%,召回率为84.2%,F1值达到84.9%,表明该方法在保障较高准确率的同时,具备良好的覆盖能力.

综上,本文提出的混合抽取框架在共伴生矿产文本处理中表现出良好的适应性与稳定性.通过融合上下文语义建模与领域知识引导策略,模型在实体识别和关系抽取任务中均体现出更高的判别精度,显著提升了抽取结果的准确性与一致性.该方法为矿产领域知识图谱的自动化构建提供了结构清晰、语义准确的数据基础,验证了知识增强型深度学习方法在地学文本处理中的应用潜力.

### 3.2 知识融合结果

本文提出的融合策略在实体识别与关系判别层面均展现出显著优化效果,特别是在处理命名异构、表达多样及语义模糊等问题上具有较强适应性.为进一步量化其提升效果,本文从知识一致性

(如实体重复率、表述标准化率)、图谱规模与结构(节点与边数量)以及构建效率(人工校验工作量)3个维度对融合前后结果进行系统评估,如表6所示.

融合后实体重复率降低5.7个百分点,标准化率提升16.3%,图谱节点与边数分别减少14.0%和6.1%,人工校验工作量下降56.2%,表明该策略在提升知识一致性的同时有效精简了图谱结构并降低了构建成本.

### 3.3 共伴生矿产知识图谱服务与智能应用

共伴生矿产具有多矿种共生、成因耦合、赋存关系复杂等典型特征,其资源评价与勘察决策高度依赖对矿物组合、空间配置及成矿协同性的系统认知.在知识图谱构建的基础上,本文围绕共伴生矿产的知识可视化、智能问答与成矿预测开展系统化应用探索.基于Neo4j提供的Cypher查询语言,设计并实现了实体检索、路径查找与子图匹配等核心查询功能,有效支持对高连接度地质知识网络的多跳检索与语义挖掘.

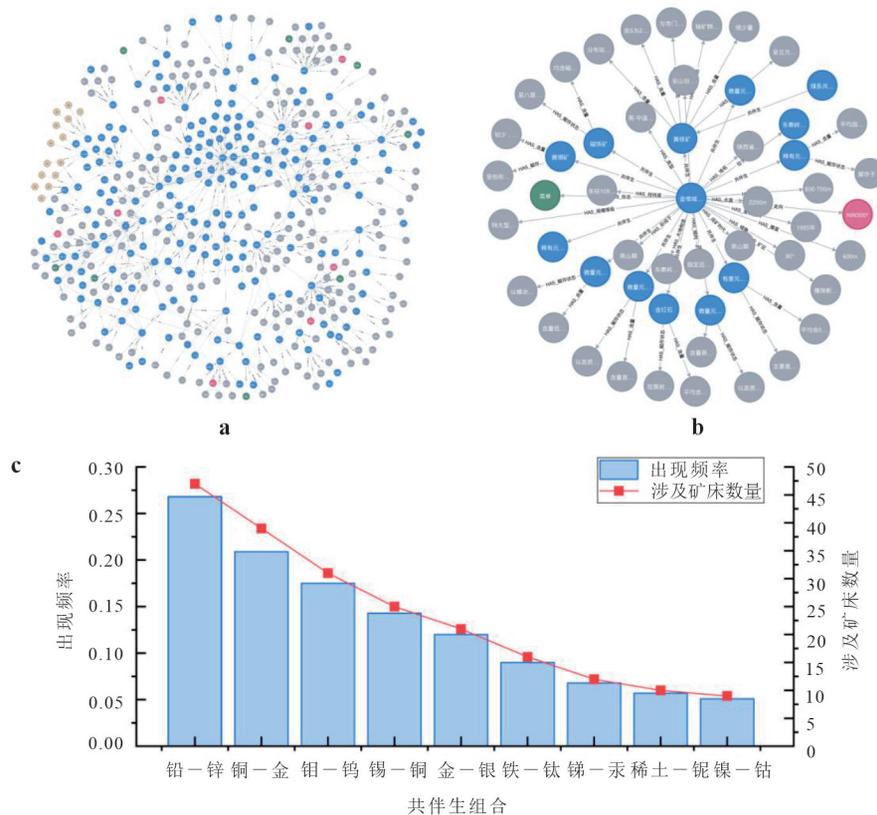


图 6 共生矿产知识图谱可视化与特征分析

Fig.6 Visualization and feature analysis of the knowledge graph for Co-associated Minerals

由于印刷限制,详细 a,b 图见电子附件

**3.3.1 知识图谱可视化与特征分析** 图 6a 展示了所构建的知识图谱的整体拓扑结构,节点代表矿区、矿床、矿物种类等核心实体,边表示“共生”、“赋存于”等地质关系,并附带诸如“规模等级”、“成矿类型”、“空间位置”等属性信息,增强了图谱的语义表达能力.截至当前版本,该知识图谱已包含了 1.2 万余个节点和 2.8 万余条关系,覆盖了多个重要成矿区带.数据主要来源于 100 余份地质勘查报告与 387 篇科研文献,确保了知识图谱具有良好的广度和代表性.

图 6b 是以“金堆城钼矿”为典型矿床构建的知识子图,集中呈现了矿床尺度的精细信息,涵盖其成因类型、区域构造背景、成矿时代及岩性特征等多个维度.该子图基于 Cypher 查询语言从知识图谱中精准抽取相关节点与关系构建而成,该方法有助于聚焦特定矿床的详细属性,有效规避了全局图结构中常见的视觉重叠与信息冗余问题.通过可视化查询界面,可清晰展示该矿床的斑岩型成因特征及其所处的燕山期岩浆热液成矿背景.此外,子图进一步揭示了除主矿种钼之外,共伴生的黄铁矿、黄

铜矿、磁铁矿等多种矿物信息,并明确了这些矿物中金、银、硒、铼等有益元素的赋存状态,为矿产资源综合评价与综合利用提供了重要参考.

图 6c 揭示了共生矿产图谱中出现频率最高的共生矿物组合,其中“铅-锌”“铜-金”“钨-钼”最为常见,广泛分布于我国多个重要成矿区带.这些组合多形成于特定的成矿环境中,如热液型、斑岩型或矽卡岩型矿床,反映出特定地球化学条件下元素的协同富集规律.例如,铅锌组合常与热液活动或沉积改造作用相关,铜金组合多见于斑岩系统,而钨钼则常见于中高温岩浆热液体系.该统计结果不仅体现了共生矿产的典型成因模式,也验证了知识图谱对区域成矿规律的有效表达能力.

**3.3.2 智能问答** 为提升知识图谱的易用性,本文开发了基于语义解析的智能问答模块(图 7).该系统采用领域自适应分词与实体识别技术,能够准确理解用户自然语言问题中的地质实体与查询意图.如图 7a 所示,系统首先通过专业地质词典增强的分词模块识别关键实体,然后基于预定义的意图分类规则识别查询类型,最终将语义解析结果转换为

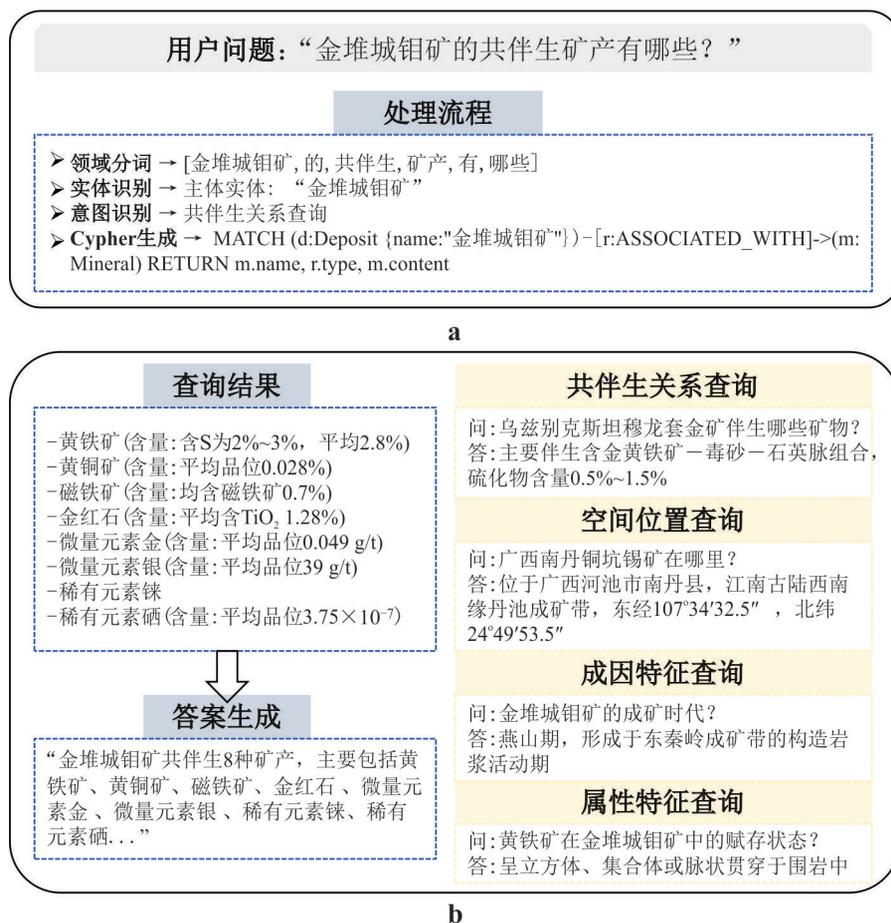


图7 智能问答系统架构与功能示例

Fig.7 Architecture and functional examples of the intelligent question-answering system

Cypher查询语句在知识图谱中执行。

系统支持多种类型的知识查询(图7b):在共伴生关系查询方面,能够准确返回特定矿床的完整伴生矿物组合及其详细的含量、赋存状态信息,如查询“金堆城钼矿的共伴生矿产”可返回黄铁矿、黄铜矿等8种主要伴生矿物及其品位数据;在空间位置查询方面,可提供矿床的大地构造位置、精确经纬度坐标及区域地质背景;在成因特征查询方面,能够识别成矿时代并关联同期区域构造岩浆事件,实现深层次的地质知识挖掘。测试结果表明,在一个由领域专家构建并审核的测试集上(共包含50个典型地质问题),系统在多类查询任务上的平均准确率达到86.3%。

**3.3.3 成矿预测** 本研究基于知识图谱中整合的矿床成矿模式信息,提出一种面向共伴生矿产的多维度相似性匹配预测方法。该方法从知识图谱中提取候选矿床的关键特征,包括矿床类型、目标矿产、成矿时代、构造背景及共伴生矿物组合等,并将其编码为结构化特征向量;在此基础上,采用改进的

余弦相似度算法,结合地质认识对不同特征赋予相应权重,计算目标区域与已知矿床之间的综合相似度。依据相似度排序结果,生成成矿潜力评估报告,为勘查靶区优选提供依据。

如图8a所示,预测流程主要包括特征提取、相似度计算与预测结论3个环节;图8b以寻找斑岩型钼矿有利靶区为例,展示了系统输出的预测分析报告,其中金堆城钼矿匹配度达95%,结合其成因类型、成矿时代与共伴生组合特征,推荐东秦岭成矿带等区域作为优先勘查区,并提出预期共伴生矿物及找矿标志,为区域矿产勘查提供了可解释的支持依据。

**3.3.4 共伴生矿产大数据平台** 为支撑共伴生矿产知识的系统化管理与可视化分析,本文构建了共伴生矿产数据大平台。如图9所示,该平台整合重点成矿区带的矿床类型、矿物组合、成矿时代及构造背景等多源信息,形成结构化数据基础,并融合知识图谱与地理信息系统,实现属性语义与空间分布的联动展示。用户可在知识图谱中探索矿产间的成

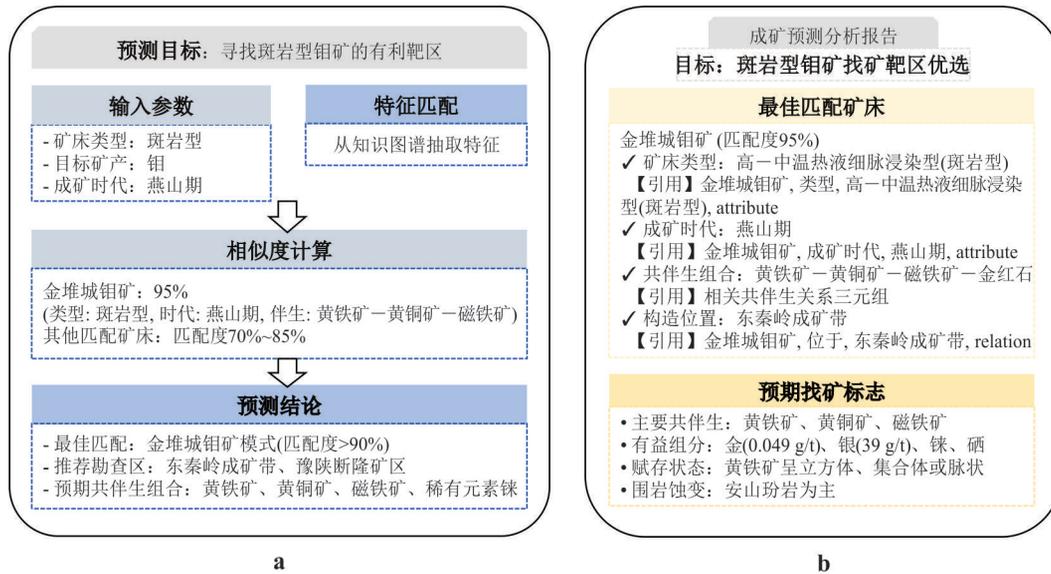


图 8 知识图谱驱动的成矿预测示例

Fig.8 Example of mineral prospectivity prediction using the co-associated mineral knowledge graph

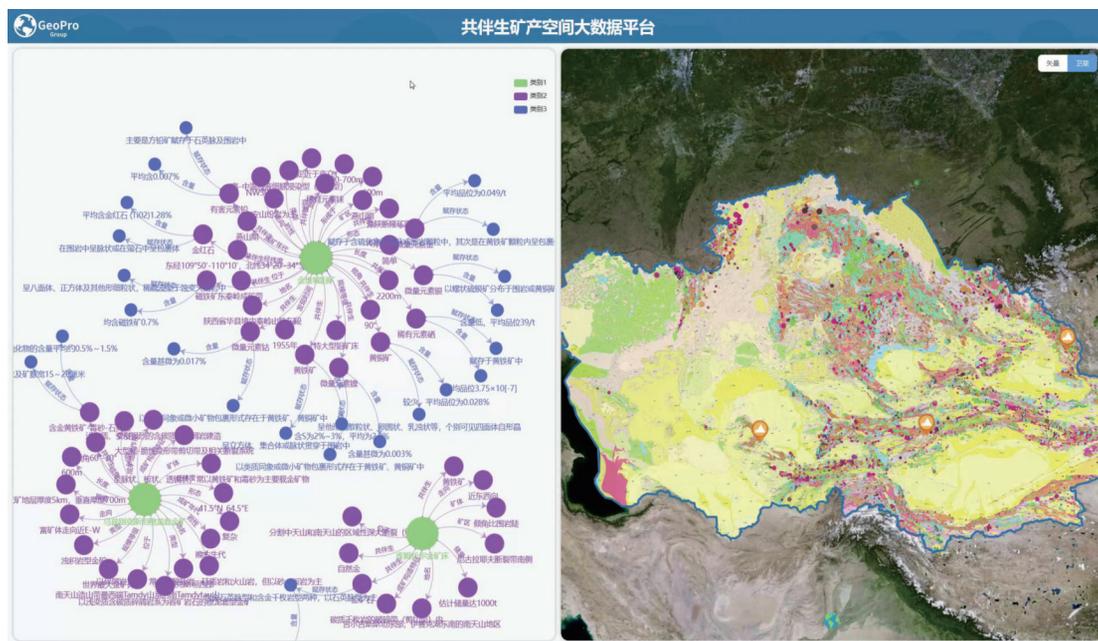


图 9 共生矿产大数据平台界面(示例)

Fig.9 Interface of the co-associated mineral data platform

因关联与共生关系,同时在地图上直观查看矿床分布与区域构造格局.平台支持双向交互:点击图谱节点可定位其在地图上的位置并获取详细地质信息,也可通过地图选区反向查询对应区域的矿产知识网络.该机制显著提升了数据的可解释性,为区域成矿规律分析与勘查方向研判提供有力支撑.

#### 4 结论与展望

本文面向共生矿产知识的结构化表达与智能应用需求,构建了数据-知识协同驱动的知识图谱方法体系,实现了从非结构化地质文本到可计算知识的高效转化.共生矿产的高度耦合性与语义隐含性,决定了其建模必须兼顾领域知识与数据规律,而本文提出的协同范式为此类复杂地学问题提供了有效路径.所形成的知识基础设施,为矿产勘

查智能化提供了可解释、可扩展的技术支撑。

(1)提出“知识引导-数据反馈”双向协同机制,破解传统模式中知识建模与数据处理割裂的问题。基于成矿理论构建初始本体,并通过多源地质数据持续迭代优化;采用融合领域词典与先验知识的增强型BERT-BiLSTM-CRF模型进行信息抽取,实体识别F1值达83.2%,较BiLSTM-CRF基线模型提升15.4个百分点,较BERT-CRF基线模型提升8.1个百分点,显著提升复杂地质术语的识别精度。

(2)构建了涵盖1.2万余个节点与2.8万余条关系的共生矿产知识图谱,集成矿区、矿床、矿物及构造单元等核心实体及其“共生”“赋存于”等地质语义关系,支持基于OWL的语义一致性校验与Neo4j的高效存储与查询。通过知识融合策略,有效解决实体歧义与语义冗余问题,图谱实体重复率降低5.7个百分点,结构显著优化,节点与边数分别减少14.0%和6.1%,在保障知识完整性的同时大幅提升图谱的紧凑性与可维护性。

(3)所构建的知识图谱已支撑多元化的知识服务应用:通过融合Neo4j与地理信息系统,实现矿产知识的可视化分析与时空关联展示;依托大数据平台,支持矿床特征的交互式查询与共生关系追溯;开发智能问答系统(准确率达86.3%),提升知识获取效率;并基于图谱模式匹配,开展成矿预测与靶区推荐。上述应用初步形成了“数据-知识-推理-服务”的过程,有效支撑矿产勘查中的多场景信息提取与辅助决策,展现出良好的实用性与推广前景。

未来研究将面向更复杂的矿产场景,进一步拓展知识图谱的动态更新与多跳推理能力,融合NLP与深度学习提升地质文本自动解析效率;同时探索跨区域、跨矿种的领域知识迁移机制,推动构建开放共享、持续演化的智慧地质知识体系。

致谢:感谢基金项目的资金支持,感谢课题组成员的协助与导师杨慧的指导。

附件见:<https://doi.org/10.3799/dqkx.2025.268>

## References

Bilenchi, I., Gramegna, F., Loseto, G., et al., 2025. Cowl: Pushing OWL 2 over the Edge. *Internet of Things*, 29: 101439. <https://doi.org/10.1016/j.iot.2024.101439>

Bull, L., Worden, K., Manson, G., et al., 2018. Active Learning for Semi-Supervised Structural Health Moni-

toring. *Journal of Sound and Vibration*, 437: 373-388. <https://doi.org/10.1016/j.jsv.2018.08.040>

- Chen, W., Wu, Y. Z., Chen, W. L., et al., 2018. Automatic Keyword Extraction Based on BiLSTM-CRF. *Computer Science*, 45(S1): 91-96, 113(in Chinese with English abstract).
- Du, M. X., Gao, B. Y., Wang, S. Z., et al., 2024. Design and Implementation of Time Metrology Vocabulary Ontology. *Electronics*, 13(14): 2828. <https://doi.org/10.3390/electronics13142828>
- E, H. H., Zhang, W. J., Xiao, S. Q., et al., 2019. Survey of Entity Relationship Extraction Based on Deep Learning. *Journal of Software*, 30(6): 1793-1818(in Chinese with English abstract).
- Han, W., Zhang, Z. K., Lei, X. Y., et al., 2025. Construction and Application of the “Rock - Mineral - Spectrum” Knowledge Graph Based on Multi-Source Fusion. *Earth Science*, 1-18[2025-11-05]. <https://link.cnki.net/urlid/42.1874.P.20251010.1336.006>(in Chinese with English abstract).
- Ju, J. H., Wan, H., Han, J., 2024. Overview of Characteristics and Comprehensive Evaluation Points of Associated Mineral Resources in China. *China Mining Magazine*, 33(12): 17-23(in Chinese with English abstract).
- Li, D. M., Zhang, Y., Li, D. Y., et al., 2020. Review of Entity Relation Extraction Methods. *Journal of Computer Research and Development*, 57(7): 1424-1448(in Chinese with English abstract).
- Li, J., Sun, A. X., Han, J. L., et al., 2022. A Survey on Deep Learning for Named Entity Recognition. *IEEE Transactions on Knowledge and Data Engineering*, 34(1): 50-70. <https://doi.org/10.1109/TKDE.2020.2981314>
- Liao, Z. X., 2021. The Construction and Application of Knowledge Graph for Geology(Dissertation). University of Electronic Science and Technology of China, Chengdu(in Chinese with English abstract).
- Meng, F. Q., Yang, S. S., Wang, J. D., et al., 2022. Creating Knowledge Graph of Electric Power Equipment Faults Based on BERT-BiLSTM-CRF Model. *Journal of Electrical Engineering & Technology*, 17(4): 2507-2516. <https://doi.org/10.1007/s42835-022-01032-3>
- Pérez, J., Díaz, J., Garcia-Martin, J., et al., 2020. Systematic Literature Reviews in Software Engineering: Enhancement of the Study Selection Process Using Cohen's Kappa Statistic. *Journal of Systems and Software*, 168: 110657. <https://doi.org/10.1016/j.jss.2020.110657>
- Qiu, Q. J., Xie, Z., Wu, L., et al., 2019. BiLSTM-CRF for Geological Named Entity Recognition from the Geoscience

- Literature. *Earth Science Informatics*, 12(4): 565–579. <https://doi.org/10.1007/s12145-019-00390-3>
- Qiu, Y. C., 2021. Design and Application of Knowledge Graph Visualization Platform(Dissertation). Beijing University of Posts and Telecommunications: Beijing(in Chinese with English abstract).
- Tian, L., Zhang, J. C., Zhang, J. H., et al., 2021. Knowledge Graph Survey: Representation, Construction, Reasoning and Knowledge Hypergraph Theory. *Journal of Computer Applications*, 41(8): 2161–2186(in Chinese with English abstract).
- Wang, C. B., Wang, M. G., Wang, B., et al., 2024. Knowledge Graph-Infused Quantitative Mineral Resource Forecasting. *Earth Science Frontiers*, 31(4): 26–36(in Chinese with English abstract).
- Wang, C., 2021. Extraction of Spatiotemporal Attributes Information of Gold Mines and Visual Expression of Knowledge Graphs for Chinese Literature(Dissertation). Hefei University of Technology, Hefei(in Chinese with English abstract).
- Wang, H., Zhang, Q. Q., Cai, W. W., et al., 2017. Research on Storage Method for Domain Ontology Based on Neo4j. *Application Research of Computers*, 34(8): 2404–2407(in Chinese with English abstract).
- Wang, N. Y., Ye, Y. X., Liu, L., et al., 2021. Language Models Based on Deep Learning: a Review. *Journal of Software*, 32(4): 1082–1115(in Chinese with English abstract).
- Wang, Q., Huang, J., Liu, X., et al., 2022. Construction and Visualization of Knowledge Graph of Geological Report. *Computer Systems and Applications*, 31(8): 140–145(in Chinese with English abstract).
- Wu, R. Z., Li, H., Mei, H. B., et al., 2025. A Landslide Monitoring and Early Warning System with Retrieval-Augmented Generation Enhanced by Knowledge Graph. *Earth Science*, 50(10): 4125–4136(in Chinese with English abstract).
- Wu, Z. W., Huang, B. X., Zhao, X. L., et al., 2024. Occurrence Characteristics and Distribution of Paragenetic Commercial Minerals and Associated Minerals in Coal-Bearing Strata in China. *Journal of Mining & Safety Engineering*, 41(1): 29–46(in Chinese with English abstract).
- Xie, T., Yang, J. A., Liu, H., 2020. Chinese Entity Recognition Based on BERT-BiLSTM-CRF Model. *Computer Systems & Applications*, 29(7): 48–55(in Chinese with English abstract).
- Xu, H. R., Xiang, Y., Ding, L., et al., 2025. Chinese Named Entity Recognition Method Based on Knowledge Enhancement from Large Language Models and Multi-feature Fusion. *Journal of Beijing University of Aeronautics and Astronautics*, 1–12 <https://doi.org/10.13700/j.bh.1001-5965.2024.0421>(in Chinese).
- Yang, H., Sheng, Y. H., Wen, Y. N., et al., 2008. Research on Standardization and Formalization of Heterogeneous Geo-Analysis Models in Web Space. *Geo-Information Science*, 10(3): 382–389(in Chinese with English abstract).
- Yang, J., 2020. Design and Implementation of Application Performance Management System Based on APM(Dissertation). Nanjing University, Nanjing(in Chinese with English abstract).
- Yang, X. Z., 2016. Research and Implementation on Entity Alignment and Attribute Alignment(Dissertation). Beijing Institute of Technology, Beijing(in Chinese with English abstract).
- Ye, Y. X., Liu, J. W., Zeng, W. X., et al., 2024. Ontology-Guided Knowledge Graph Construction for Mineral Prediction. *Earth Science Frontiers*, 31(4): 16–25(in Chinese with English abstract).
- Yin, Q., Zhou, Y., Rao, J. Y., 2025. An Intelligent Prediction Method for Surface Settlement of Shield Tunnel Construction Based on CNN-BiLSTM-SA. *KSCE Journal of Civil Engineering*, 29(4): 100061. <https://doi.org/10.1016/j.kscej.2024.100061>
- Yu, F., Wang, M. G., Wang, C. B., et al., 2025. GeoMinLM: a Large Language Model in Geology and Mineral Survey in Yunnan Province. *Ore Geology Reviews*, 182: 106638. <https://doi.org/10.1016/j.oregeorev.2025.106638>
- Zhang, B. Y., Tang, J. C., Zhang, T. Y., et al., 2025. Knowledge Graph and Question-Answering Model for Geological Prospecting Empowered by Large Language Models. *Earth Science*. 1–24[2025-10-10].<https://doi.org/10.3799/dqkx.2025.176>(in Chinese with English abstract).
- Zhang, L., Lobov, A., 2024. Semantic Web Rule Language-Based Approach for Implementing Knowledge-Based Engineering Systems. *Advanced Engineering Informatics*, 62: 102587. <https://doi.org/10.1016/j.aei.2024.102587>
- Zhang, Q. L., Zhou, Y. Z., Guo, L. X., et al., 2024. Intelligent Application of Knowledge Graphs in Mineral Prospecting: a Case Study of Porphyry Copper Deposits in the Qin-Hang Metallogenic Belt. *Earth Science Frontiers*, 31(4): 7–15(in Chinese with English abstract).
- Zhao, P. D., 2019. Characteristics and Rational Utilization of Geological Big Data. *Earth Science Frontiers*, 26(4): 1–5 (in Chinese with English abstract).
- Zhao, T., Liu, C., Li, H. M., et al., 2025. Innovative Applications of Data Mining and Knowledge Graph Construction

- in National Mineral Resources Survey. *Earth Science Frontiers*, 1–14[2025-07-16].<https://doi.org/10.13745/j.esf.sf.2024.11.66>(in Chinese with English abstract).
- Zhou, C. H., Wang, H., Wang, C. S., et al., 2021. Geoscience Knowledge Graph in the Big Data Era. *Science China Earth Sciences*, 64(7): 1105–1114. <https://doi.org/10.1007/s11430-020-9750-4>
- 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 with English abstract).
- Zhou, Y. Z., Xiao, F., 2024. Overview: a Glimpse of the Latest Advances in Artificial Intelligence and Big Data Geoscience Research. *Earth Science Frontiers*, 31(4): 1–6(in Chinese with English abstract).
- Zhou, Y. Z., Zhang, Q. L., Huang, Y. J., et al., 2021. Constructing Knowledge Graph for the Porphyry Copper Deposit in the Qingzhou-Hangzhou Bay Area: Insight into Knowledge Graph Based Mineral Resource Prediction and Evaluation. *Earth Science Frontiers*, 28(3): 67–75(in Chinese with English abstract).
- Zhuang, Z. H., Jiao, S. L., Sun, C. H., 2024. Named Entity Recognition in Geological Field Based on BERT-BiLSTM-CRF. *Computer & Digital Engineering*, 52(6): 1815–1820, 1876(in Chinese with English abstract).
- Zuo, R. G., 2021. Data Science-Based Theory and Method of Quantitative Prediction of Mineral Resources. *Earth Science Frontiers*, 28(3): 49–55(in Chinese with English abstract).
- ### 中文参考文献
- 陈伟, 吴友政, 陈文亮, 等, 2018. 基于BiLSTM-CRF的关键词自动抽取. *计算机科学*, 45(S1): 91–96, 113.
- 鄂海红, 张文静, 肖思琪, 等, 2019. 深度学习实体关系抽取研究综述. *软件学报*, 30(6): 1793–1818.
- 韩伟, 张照坤, 雷辛亚, 等, 2025. 多源融合的“岩石—矿物—光谱”知识图谱构建及应用. *地球科学*, 1–18[2025-11-05]. <https://link.cnki.net/urlid/42.1874.P.20251010.1336.006>
- 鞠建华, 万会, 韩见, 2024. 我国共生矿产资源特点及综合评价要点综述. *中国矿业*, 33(12): 17–23.
- 李冬梅, 张扬, 李东远, 等, 2020. 实体关系抽取方法研究综述. *计算机研究与发展*, 57(7): 1424–1448.
- 廖振鑫, 2021. 面向地质领域的知识图谱构建研究及应用(硕士学位论文). 成都: 电子科技大学.
- 邱宇辰, 2021. 知识图谱可视化平台的设计实现及应用(硕士学位论文). 北京: 北京邮电大学.
- 田玲, 张谨川, 张晋豪, 等, 2021. 知识图谱综述: 表示、构建、推理与知识超图理论. *计算机应用*, 41(8): 2161–2186.
- 汪陈, 2021. 面向中文文献的金矿时空属性信息抽取及知识图谱可视化表达(硕士学位论文). 合肥: 合肥工业大学.
- 王成彬, 王明果, 王博, 等, 2024. 融合知识图谱的矿产资源定量预测. *地学前缘*, 31(4): 26–36.
- 王红, 张青青, 蔡伟伟, 等, 2017. 基于Neo4j的领域本体存储方法研究. *计算机应用研究*, 34(8): 2404–2407.
- 王乃钰, 叶育鑫, 刘露, 等, 2021. 基于深度学习的语言模型研究进展. *软件学报*, 32(4): 1082–1115.
- 王晴, 黄进, 刘鑫, 等, 2022. 成果地质资料知识图谱构建与可视化. *计算机系统应用*, 31(8): 140–145.
- 吴润泽, 李浩, 梅红波, 等, 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.
- 吴占伟, 黄炳香, 赵兴龙, 等, 2024. 我国煤系共生矿产赋存特征及分布规律. *采矿与安全工程学报*, 41(1): 29–46.
- 谢腾, 杨俊安, 刘辉, 2020. 基于BERT-BiLSTM-CRF模型的中文实体识别. *计算机系统应用*, 29(7): 48–55.
- 徐浩然, 向阳, 丁玲, 2025. 基于大语言模型知识增强和多特征融合的中文命名实体识别方法. *北京航空航天大学学报*, 1–12[2025-06-28].<https://doi.org/10.13700/j.bh.1001-5965.2024.0421>.
- 杨军, 2020. 基于APM的应用性能管理系统的设计与实现(硕士学位论文). 南京: 南京大学.
- 杨秀璋, 2016. 实体和属性对齐方法的研究与实现(硕士学位论文). 北京: 北京理工大学.
- 叶育鑫, 刘家文, 曾婉馨, 等, 2024. 基于本体指导的矿产预测知识图谱构建研究. *地学前缘*, 31(4): 16–25.
- 张宝一, 唐嘉成, 张彤蕴, 等, 2025. 大语言模型赋能的地质找矿知识图谱与问答模型构建. *地球科学*, 1–24[2025-10-10]. <https://doi.org/10.3799/dqkx.2025.176>.
- 张前龙, 周永章, 郭兰萱, 等, 2024. 找矿知识图谱的智能化应用: 以钦杭成矿带斑岩铜矿为例. *地学前缘*, 31(4): 7–15.
- 赵鹏大, 2019. 地质大数据特点及其合理开发利用. *地学前缘*, 26(4): 1–5.
- 赵汀, 刘超, 李厚民, 等, 2025. 矿产资源国情调查中数据挖掘创新应用与知识图谱构建. *地学前缘*, 1–14[2025-07-16]. <https://doi.org/10.13745/j.esf.sf.2024.11.66>.
- 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.
- 周永章, 肖凡, 2024. 管窥人工智能与大数据地球科学研究新进展. *地学前缘*, 31(4): 1–6.
- 周永章, 张前龙, 黄永健, 等, 2021. 钦杭成矿带斑岩铜矿知识图谱构建及应用展望. *地学前缘*, 28(3): 67–75.