

# 基于大语言模型的华南稀有金属矿床知识图谱构建

吴海陆<sup>1</sup>, 窦磊<sup>1\*</sup>, 虞鹏鹏<sup>2\*</sup>, 朱世博<sup>1</sup>, 余德延<sup>1</sup>

1. 广东省地质调查研究院, 广东广州 510110

2. 中山大学地球科学与工程学院, 广东省地质过程与矿产资源探查重点实验室, 广东珠海 519082

**摘要:** 以大语言模型为代表的新一代人工智能技术为地学知识的结构化表达与智能推理提供了新机遇。针对地质领域知识体系复杂, 非结构化文本语义分散、难以再利用和可视化的问题, 本文以华南稀有金属矿床为研究对象, 提出了一种融合矿床成因与找矿标志的统一知识图谱构建策略。研究基于 Deepseek R1-32B 大语言模型与提示词工程, 从大量地质文献中自动抽取并构建了涵盖 Li、Be、Nb、Ta 等关键稀有矿种的知识图谱。从知识图谱及其拓展性分析的结果表明, 华南稀有金属成矿与印支期、燕山期岩浆活动密切相关, 具有显著的高分异与岩浆热液作用特征; 稀有金属元素呈现 Li-Be-Nb-Ta-W-Sn 的组合异常。综上所述, 基于大语言模型构建的知识图谱揭示了华南稀有金属成矿的多阶段成矿机制, 阐明了稀有金属矿床在地球化学异常、构造控制及蚀变分带方面的内在联系, 为华南及邻区的稀有金属勘查提供了智能化研究方案。

**关键词:** 知识图谱; 大语言模型; 稀有金属成矿; 智能勘查; 华南

中图分类号: TP391.1 收稿日期: 2025-11-26

## Construction of a Knowledge Graph for Rare Metal Deposits in South China Based on Large Language Models

基金项目: 广东省重点矿种找矿靶区优选及找矿目标定位(任务书编号: 2024-47); 广东省自然科学基金项目(No.2024A1515030216)。

第一作者: 吴海陆(1984-), 男, 高级工程师, 主要从事矿产地质勘查、数字地质调查, Email: wuhailu77887@163.com;

\*通讯作者: 窦磊(1979-), 男, 教授级高级工程师, 主要从事地球化学, Email: ggsdl@163.com;

虞鹏鹏(1991-), 男, 副教授, 博士生导师, 主要从事矿产勘查与成因研究, e-mail: yupp3@mail.sysu.edu.cn

**Abstract:**The new generation of artificial intelligence technologies, represented by large language models (LLMs), provides new opportunities for the structured representation and intelligent reasoning of geological knowledge. To address the challenges posed by the complexity of geoscientific knowledge systems, as well as the semantic fragmentation, limited reusability, and poor visualizability of unstructured texts, this study focuses on rare metal deposits in South China and proposes a unified strategy for constructing a knowledge graph that integrates deposit genesis and prospecting indicators. Based on the Deepseek R1-32B large language model and prompt engineering, the study automatically extracts and constructs a knowledge graph covering key rare metal elements such as Li, Be, Nb, Ta. Based on the knowledge graph construction and its extensibility analysis, the results indicate that rare metal mineralization in South China is closely associated with Indosinian and Yanshanian magmatic activities, characterized by significant high-degree fractionation and magmatic-hydrothermal processes. Rare metal elements exhibit a combinatorial anomaly of Li–Be–Nb–Ta–W–Sn. In summary, the knowledge graph constructed using LLMs reveals the multi-stage metallogenic mechanisms of rare metals in South China, clarifies the intrinsic relationships among geochemical anomalies, structural controls, and alteration zoning of rare metal deposits, and provides an intelligent research framework for the exploration of rare metals in South China and adjacent regions.

**Keywords:** Knowledge graph; Large language model; Rare metal mineralization; Intelligent exploration; South China

## 引言

华南地区位于扬子地块与华夏地块的结合带上,是中国稀有金属成矿最为集中的地区之一(李晓峰等, 2021; 吴福元等, 2023; 赵正等, 2022)。区域内存在以钦州-萍乡-江山-杭州深大断裂(简称钦-杭大断裂)为主干的多条 NE 向与近 EW 向断裂(吴堃虹等, 2023)。在空间上,稀有金属矿床沿钦-杭断裂带两侧及其分支呈带状展布,形成南岭成矿带、赣东-闽西成矿带和东南沿海火山岩成矿带等多个富集区(秦锦华等, 2023)。主要矿床类型包括花岗岩型、伟晶岩型、云英岩/石英脉型、接触交代型与火山岩型。典型矿床如江西宜春 414、栗木、甘坊、松树岗、银山等,均富集 Li-Be-Nb-Ta-Sn-Rb 等元素(侯占德等, 2023; 吴堃虹等, 2023),反映出华南地区典型的高分异花岗岩-伟晶岩体系特征(Xiao et al., 2025)。学界针对华南稀有金属的研究较为深入,累计了海量的地质、地球化学、地球物理等结构化数据以及文本等非结构化数据。其中,文本数据蕴含了丰富的成矿和找矿信息,对其的知识获取往往需要耗费大量的时间阅读文献;更为重要的是,对其的分析依赖深厚的专业背景和科研能力。大数据和人工智能技术的发展为探究地质矿产领域文本数据处理、知识库建设与可视化提供了可靠途径(Guan et al., 2019)。

数学地球科学是一门融合数学、计算机科学与地球科学的交叉学科(成秋明, 2021)。该领域的研究主要围绕岩石形成过程中的地质数据(如岩石参数、地球化学数据、地球物理勘探数据和卫星影像等)获取、定量分析和新知识挖掘(Cracknell and Reading, 2014; Lima et al., 2017; Wang et al., 2018; Xiao et al., 2016)。近年来,大数据与人工智能算法的引入开启了地质学的智能化时代(周永章等, 2021)。对于矿床而言,除结构化数据外,大量关于成矿环境与矿床特征的重要信息蕴藏于文本描述之中,这类数据难以被量化利用(Qiu et al., 2023)。随着机器学习的发展,利用文本挖掘研究地质问题的案例日益增多。例如有学者通过加拿大、美国和澳大利亚的地球科学文本数据挖掘,研究了密西西比河谷型铅锌矿床和碎屑岩主导型铅锌矿床的成矿潜力预测(Lawley et al., 2023);也有人基于深度学习与自然语言处理技术从矿产勘查报告中提取表格信息,表明文本数据挖掘能有效应用于地质学研究(Dong et al., 2023)。最成功的案例是地学知

识图谱的研究，从古生物研究、特殊矿床、元素组合到矿产资源预测均有印证 (Deng et al., 2024; Zhang et al., 2025; Zhao et al., 2025)。

稀有金属 (Li、Be、Nb、Ta 等) 在科技发展和国家战略安全方面极其重要 (Ma et al., 2024; 侯占德等, 2023; 吴堃虹等, 2023)，构建华南稀有金属矿床的可视化知识库，有助于后续的矿产资源预测和矿床成因探究。然而知识图谱的构建存在实体标注量巨大、本体模型泛化、抽取方法不灵活等问题。为此，本文提出了一种兼具矿床内部因素 (如成矿系列、成矿模型与成矿规律) 和矿床外部异常表现 (如地球化学异常、地球物理异常与遥感异常) 的矿床知识本体模型 (张振杰等, 2021)。其中，矿床内部因素包括出露地层、成矿年龄、构造信息、流体信息、矿化类型、矿石类型及岩浆信息；外部因素包括地球化学异常元素与地球物理异常。结合大语言模型 (Deepseek R1) 和提示词工程，将两者创建为知识图谱，即矿床的“成矿图谱”和“找矿图谱”，随后将其合并为矿床的统一“知识模型” (图 1)。

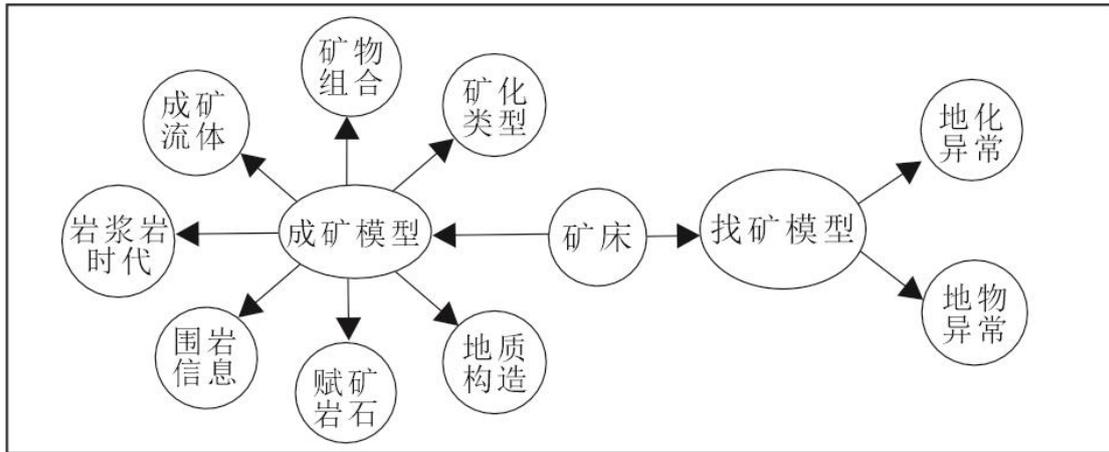


图 1. 矿床知识图谱模型结构图

Figure 1. Structure of the Mineral Deposit Knowledge Graph Model.

## 1 数据和方法

### (1) 数据来源

本研究以“华南+稀有金属”为检索词，在中国知网收集了从 2010~2025 年共 15 年的学术期刊论文作为原始数据。通过 python 工具包 fitz，提取了文献中的文本内容并做文本清洗，以 500 个字符长度为限，划分为多个可编辑的长文本片

段。

## (2) 矿床知识本体构建

承前所述，矿床知识体系要同时兼备矿床形成的内部因素（如成矿系列、成矿模型与成矿规律）和矿床形成后的外部异常表现（如地球化学异常、地球物理异常、遥感异常）。本文参考通用矿床矿产预测理论（Yan et al., 2023; 朱裕生, 2006; 朱章森和朱磊, 1998），确定了“成矿”与“找矿”两个本体模型，并为其创建了唯一的实体层 ID（对应括号中的数字），来实现后续的实体归类和图谱创建。具体的实体案例如表 1 所示。

表 1. 矿床知识本体模型信息表

Table 1. Mineral Deposit Knowledge Ontology Model Information Table

Table 1. Mineral Deposit Knowledge Ontology Model Information Table		
成矿本体因子 (0)	地层信息 (1)	地层 1 地层 2
	时代信息 (2)	区域断裂形成时代
		成矿岩体的成岩年龄
		岩石时代
		岩浆侵入时间
	地质构造 (3)	成矿时代
		区域构造带
		深大断裂
		次级断裂
		深大断裂数量
小断裂数量		
流体信息 (4)	深断裂走向	
	小断裂走向	
	多数断裂与邻近断裂之间的夹角	
	背斜	
矿化类型 (5)	向斜	
	流体类型 1	
	流体类型 2	
含矿岩石类型 (6)	流体富集元素	
	流体温度	
矿物类型 (7)	流体性质	
	矿化类型 1	
	矿化类型 2	
	矿化类型 3	
	含矿岩石 1	
	含矿岩石 2	
	矿物类型 1	
	矿物类型 2	

	矿物组合 (8)	矿物组合 1 矿物组合 2
	岩浆信息 (9)	岩浆类型 形成岩石 岩浆成因
找矿本体因子 (00)	地球物理异常 (10)	重力异常 航磁异常
	地球化学异常 (11)	正异常元素 负异常元素

### (3) 知识抽取

命名实体是内容中信息元素的统称, 而识别信息元素并对其进行分类的任务称为命名实体识别(NER), 知识抽取是命名实体识别的最终目的(Marrero et al., 2013)。其中, 识别任务通常包括命名实体检测(NED)与命名实体分类(NEC)(Li et al., 2009)。Bikel et al. (1998) 首次采用隐马尔可夫模型(HMM)对英语地名、机构名进行识别。Ma 等提出的基于双向长短期记忆网络(BiLSTM)结合卷积神经网络(CNN)与条件随机场(CRF)的端到端序列标注命名实体识别方法, 显著提升了识别准确率(Ma and Hovy, 2016)。2019年, 谷歌提出基于Transformer的 Bert 预训练词向量生成模型, 其特征提取过程考虑了上下文信息(Devlin et al., 2019)。后来, 基于 Bert-BiLSTM-CRF 的命名实体识别在识别准确率和可操作性上有了显著的提升。但是以上方法都存在大量的文本标注工作, 同时实体库和关系库的构建依赖于强大的专业背景, 在处理一词多义和个人用语习惯上存在很大缺陷。大语言模型的出现, 在语义理解和语句生成方面实现了质的飞跃。提示词工程是现在垂直领域大语言模型构建和智能体搭建最重要的一个方式。为了让基础大语言模型理解矿床的知识框架和需要抽取的实体类型, 我们提出了基于 Deepseek R1 和提示词工程的知识抽取方法(图 2)。在提示词工程中, 我们将矿床本体上传, 让基础大语言模型按照本体结构和具体实体进行抽取, 对于每一篇论文的不同片段设置相同的 id, 最后再根据 id 将信息融合。

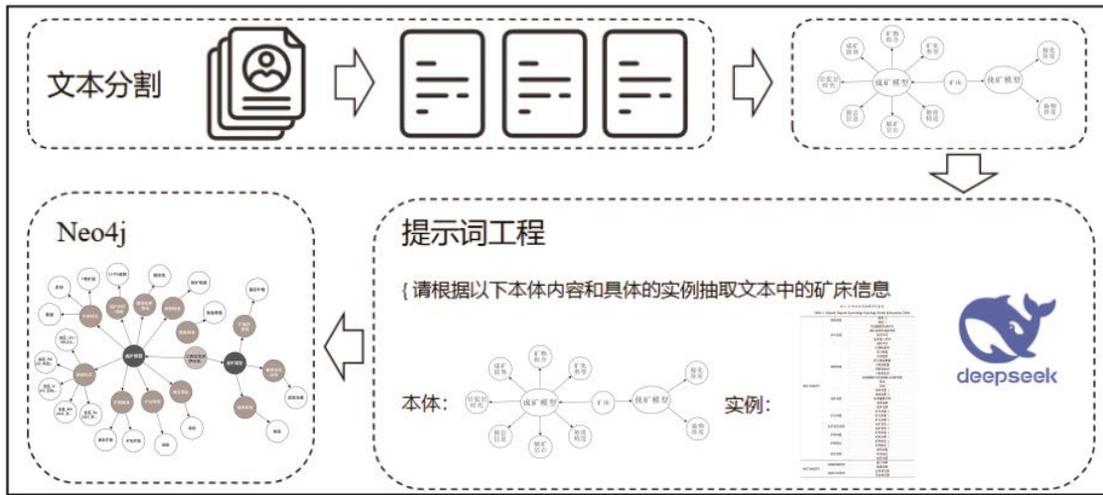


图 2. 知识图谱抽取技术路线图

Figure 2. Technical Roadmap for Knowledge Graph Extraction

#### (4) 模型测试

本研究的抽取认为主要基于提示词工程完成。首先，我们将通过 python 处理的文本段依次投喂给特定提示词工程（图 3）+ Deepseek R1-32B，Deepseek R1-32B 和在线版的全参数 Deepseek+特定提示词工程，随机抽取了 100 个语料段做性能测试，根据提前专家标记的三元组结果和实际抽取的结果，分别计算了整体抽取的准确率（Precision）、召回率（Recall）和 F1 值（图 3），从抽取结果可以看出，Deepseek R1-32B 模型本身对于三元组抽取具有一定的能力，但整体能力偏低，而通过设计的提示词工程之后，准确率（Precision）、召回率（Recall）和 F1 值都有了显著的提升，但是很明显模型在三元组抽取方面的能力依旧较弱。为此，我们又将提示词工程应用于在线版的全参数 Deep seek R，相比于 Deepseek R1-32B 很明显全参数具有显著优势，这表明 Deepseek R1 在三元组抽取任务中依然较弱，后续的研究可结合小模型微调，能使小参数模型在单项任务中赶超全参数模型。虽然本次抽取任务中 Deepseek R1 全参数模型+提示词工程具有最高的模型属性，但我们最终选择了本地部署的 Deepseek R1-32B 模型方案，因为这在一线单位的生产安全问题上具有不可替代性，另外在实际的知识图谱上，我们发现当前模型抽取的图谱已具备较好的问题说明能力。

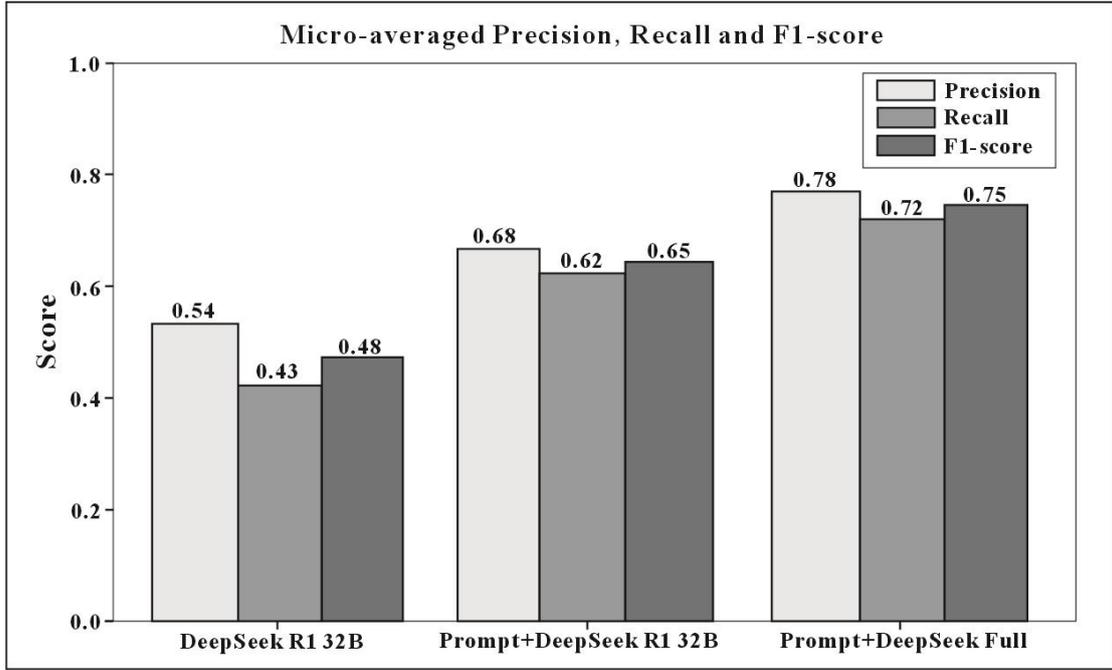


图 3. 100 个案例在不同模型下的参数对比图

Figure 3. Comparison of model parameters across 100 cases under different models.

#### (5) 知识图谱可视化

Neo4j 是常用的图数据库之一，通过 Python 语言实现数据库的连接和访问。根据大语言模型抽取的三元组数据，我们通过人工核查确认了实体类别，根据实体层 ID，创建了层级别间的包含关系，比如我们设定矿床名和模型类型之间存在唯一的 Key，这样就严格遵守了不同级别节点间的网络层级关系。为了观察华南各矿床间的相关性，在最后一层节点之间，按照语义进行了融合。具体方法为通过 Node2Vec 算法 (Grover et al., 2016) (图 4) 将知识图谱的节点转换为 N 维向量，引入余弦相似度算法计算节点语义相似性。其中，余弦相似度是计算两个向量之间角度的余弦值来反映它们的相似性 (Lahitani et al., 2016)。实现过程为：假设有两个相同维度 (N×1) 的不同向量，如  $X_T = \{x_1, x_2, x_3, \dots, x_N\}$ 、 $Y_T = \{y_1, y_2, y_3, \dots, y_N\}$ ， $X_T$  与  $Y_T$  之间的余弦相似度 (式 1) 定义如下：

$$CS(X, Y) = \frac{\sum_{i=1}^N x_i * y_i}{\sqrt{\sum_{i=1}^N x_i^2} * \sqrt{\sum_{i=1}^N y_i^2}} \quad (1)$$

它的取值范围为[-1,1],具体的含义是，当两向量方向完全相反时，C 的值等

于-1, 相同时,  $C$  的值等于 1, 两向量越接近,  $C$  值越近似 1。本次设置阈值为 0.85, 共计融合 408 个节点, 实现了部分节点间的语义消歧。最后实现了华南稀有矿床的知识图谱可视化。

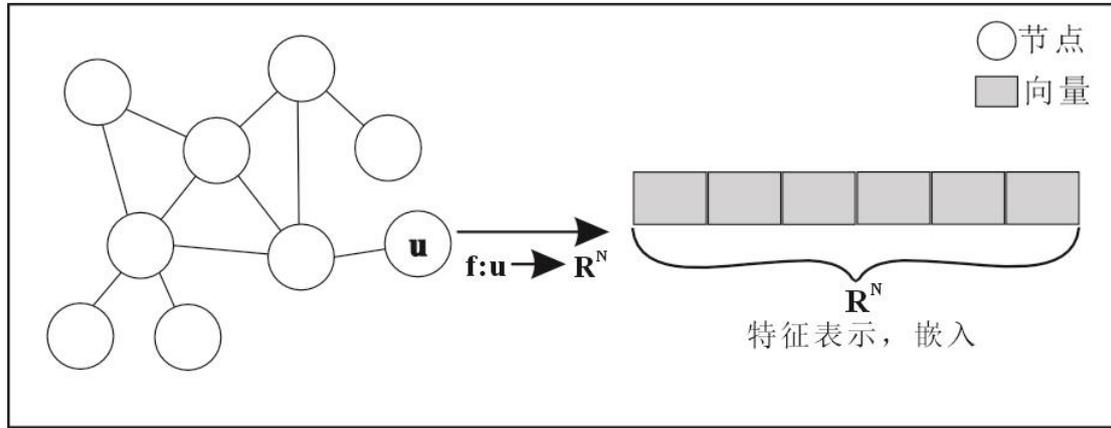


图 4. Node2vec 算法示意图

Figure 4. Schematic illustration of the Node2vec algorithm.

## 2 结果与讨论

针对华南稀有金属矿床的知识图谱, 本次抽取的矿床/矿区/矿点/岩体共计 94 个, 全图共有 5064 个节点, 5900 条关系。从整体上看节点拓展后矿床间的连接关系, 华南稀有金属矿床的研究分为两个端元, 一类是研究较为成熟的 (矿床间的连接关系紧密), 且在研究方向和结论上有一定的共识。另一部分研究可能较少, 在连接关系上以孤立或连接较少为特征。每个矿床都包含“成矿”和“找矿”两个子图 (图 5), 每下一级则是具体的知识分支, 值得注意的是, 由于大语言模型生成过程中的约束性不确定 (Maynez et al., 2020; 赵月等, 2024), 所以在总体上保持和建立本体一致的基础上, 可能会产生新的知识节点, 比如图三中的成矿知识图谱的地球化学分支, 但是这与找矿知识图谱的地球化学异常分支不同, 从子节点可以看出, 这里主要描述的应该是岩浆或者岩浆岩本身化学组成, 而非外在勘查异常。

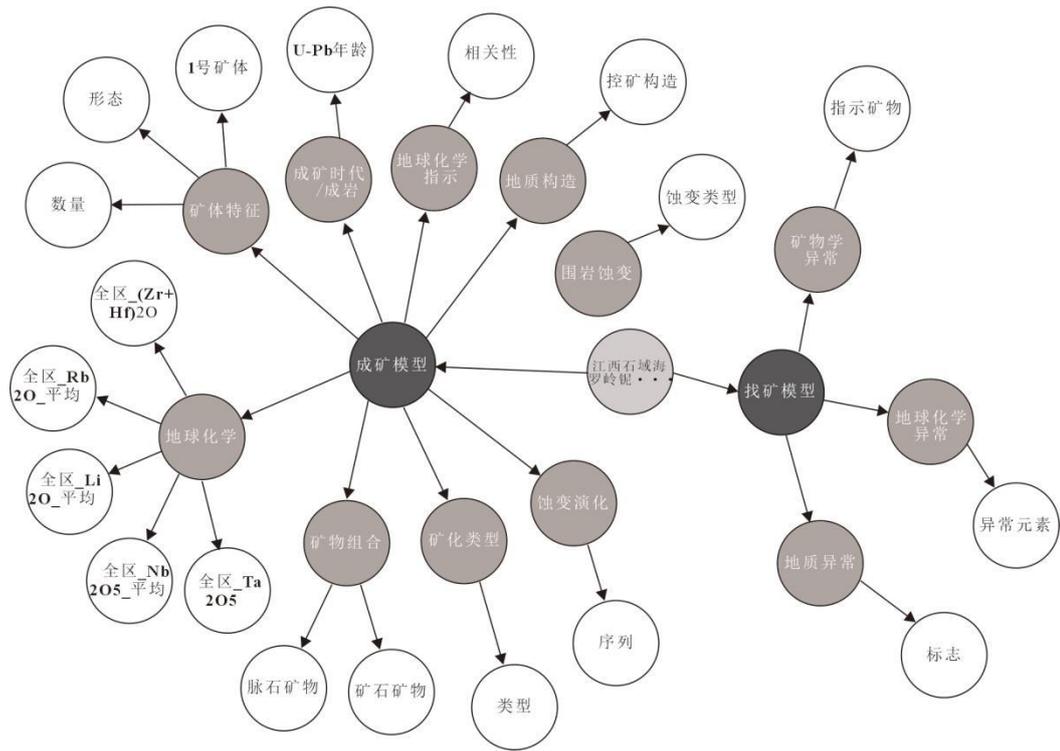


图 5. 华南稀有金属矿床（点）连接关系图

Figure 5. Connectivity Diagram of Rare Metal Deposits and Occurrences in South China

(1) 华南稀有金属矿床地球化学特征

图 6 结果揭示了华南稀有金属矿床在地球化学特征上呈现出显著的多元素耦合与分异规律。整体上，稀有金属元素在图谱中形成了以 W-Sn-Li-Be-Rb-Nb-Ta-Cs 为异常组合的地球化学聚类团（图 6a 右下），这和宏观上华南稀有金属通常与钨锡矿伴生相一致。从主量元素特征来看，华南稀有金属矿床具有高硅、低钙、过铝质的特征（图 6a 左下），微量元素上具有稀土含量较高、重稀土富集、Eu 负异常的特征（图 6a 中部和图 6b）。这些地球化学行为显示，华南稀有金属成矿可能与强烈分异的碱性花岗岩及其蚀变岩系统相关（Ou et al., 2025）。因为在高度结晶分异作用过程中，花岗质岩浆将逐渐向不相容元素聚集的熔体和流体方向演化，稀土元素则偏向于岩浆演化晚期富集，从而导致其 Nb/Ta、Zr/Hf 和 LREE/HREE(轻/重稀土)比值降低（吴福元等, 2023）。



and Occurrences in South China (a. Element Aggregation Map of Geochemical Anomalies; b. Diagram of Element Differentiation and Partition Characteristics)

(2) 华南稀有金属矿床成矿花岗岩和成矿流体特征

单从地球化学知识图谱的结果对于华南稀有金属成矿的解释具有局限性。一是缺少重要的时间线索，二是不知的岩浆的本身性质。因此我们进一步观察花岗岩和成矿流体的知识图谱（图 7），从成矿时间来看，华南稀有金属矿床的成矿和印支期、燕山期岩浆活动密切相关，具体的成矿时间分为 247~202Ma 和 160~136Ma（图 8），赋矿岩石有花岗岩、二长花岗岩、花岗斑岩、钠长石花岗岩、伟晶岩、花岗闪长岩、云英岩、黑云母花岗岩、白云母花岗岩、二云母花岗岩和锂云母花岗岩（图 9）。涉及的成矿作用主要有岩浆热液转换和热液交代，其中岩浆来源同时有幔源和壳源的参与，A 型花岗岩的结晶分离和高分异可能是稀有金属富集的主要因素。而在前人的研究中表明，华南钨锡矿虽然和稀有金属在空间上存在高度的一致性，时常伴生成矿，但是钨锡和稀有金属的成因可能存在不同（Ma et al., 2024; Yang et al., 2025; 吴福元等, 2023; 马星华等, 2025），稀有金属是岩浆高度分异的结果，而钨锡则可能是岩浆来源的热液成因（Ma et al., 2024; Song et al., 2024; 张睿泽和左仁广, 2025），这在知识图谱中具有明显的指示，因为在和岩浆相关的知识图谱中除了和高分异岩浆有关，还出现了热液交代和成矿流体的特征，其中，成矿流体具有明显的富挥发分特征，尤其是富集氟元素（图 10）。也有研究表明，挥发分中 F 和 Cl 含量的不同（图 11），恰好是控制钨锡矿中钨和锡谁占主导地位的关键（Ma et al., 2024; Yang et al., 2025）。总之，华南稀有金属成矿受燕山期和印支期岩浆活动、岩浆演化和岩浆热液的共同影响，可能由于岩浆活动的多期性，稀有金属成矿也具有多期次成矿的特征。知识图谱的构建，能有效查询和启示华南稀有金属成矿事件。

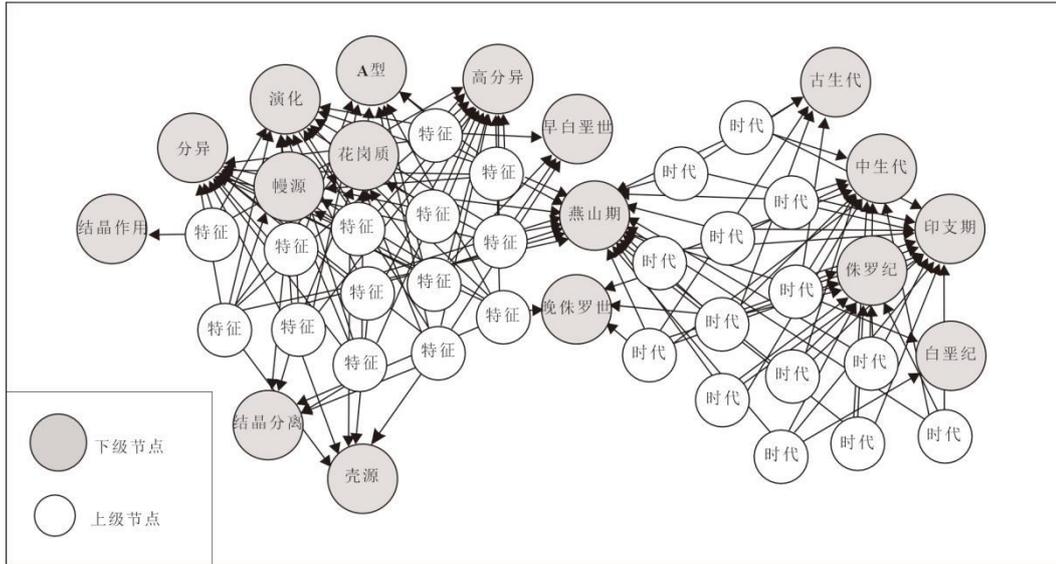


图 7. 华南稀有金属矿床（点）岩浆及其形成时代知识图谱

Figure 7. Knowledge Graph of Magmatism and Formation Epochs of Rare Metal Deposits and Occurrences in South China

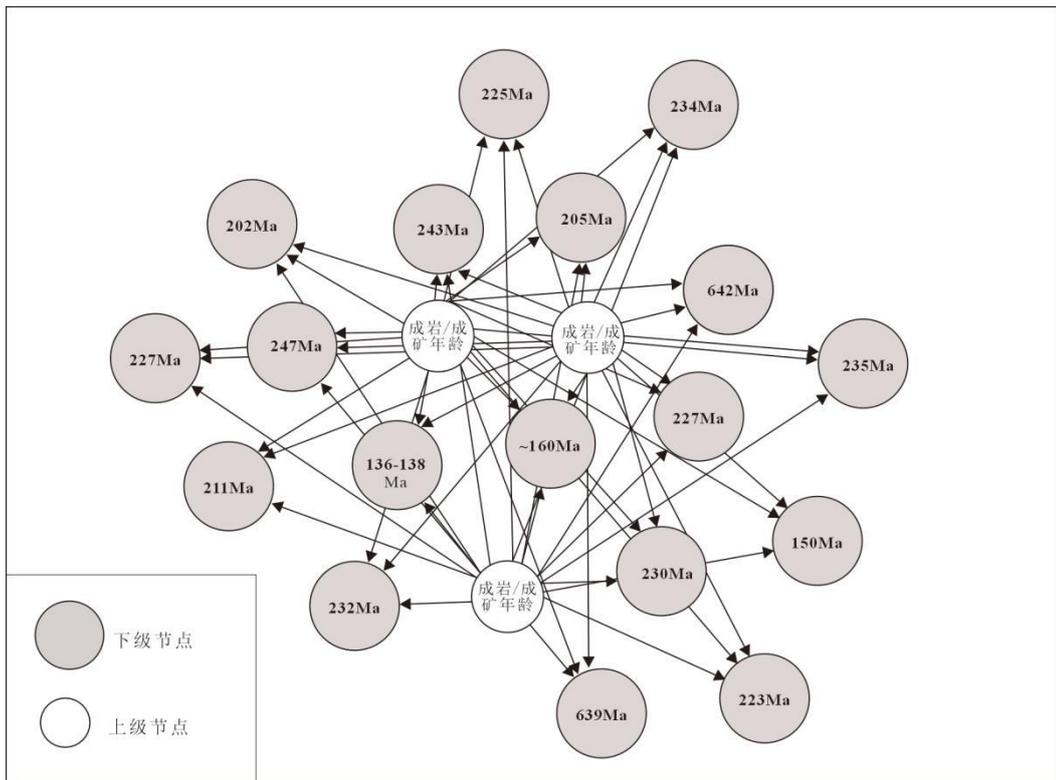


图 8. 华南稀有金属矿床（点）赋矿岩石知识图谱

Figure 8. Knowledge Graph of Host Rocks for Rare Metal Deposits and Occurrences in South China

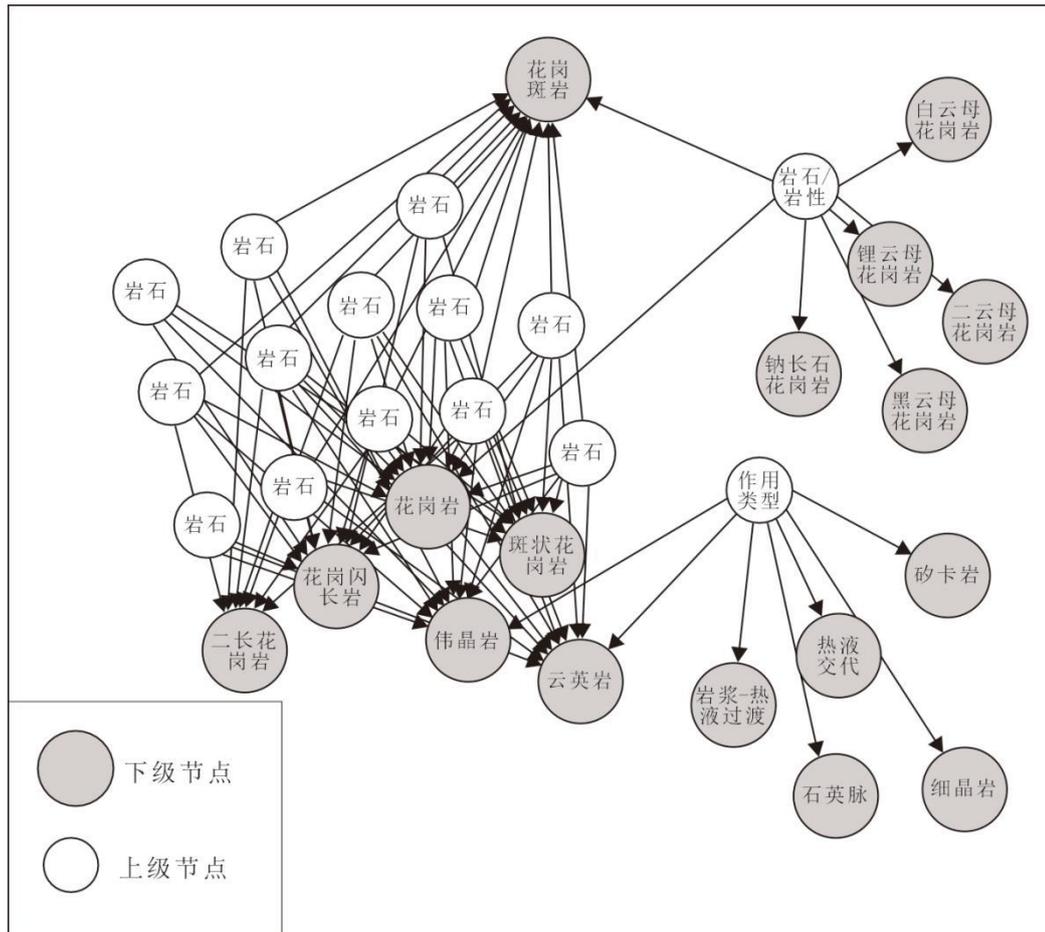


图 9. 华南稀有金属矿床（点）赋矿岩石知识图谱

Figure 9. Knowledge Graph of Host Rocks for Rare Metal Deposits and Occurrences in South China

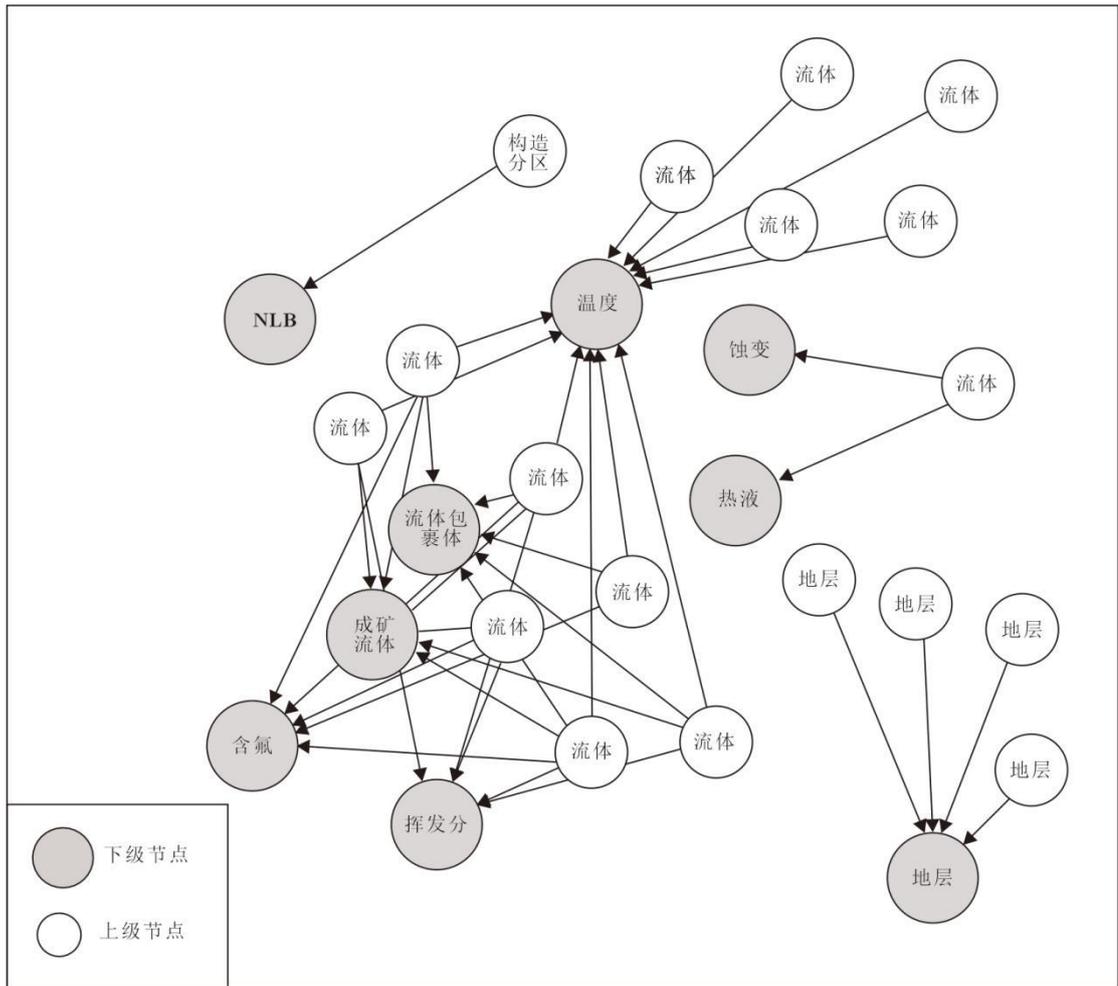


图 10. 华南稀有金属矿床（点）成矿流体特征知识图谱

Figure 10. Knowledge Graph of Ore-Forming Fluid Characteristics in Rare Metal Deposits and Occurrences in South China

### (3) 华南稀有金属矿床构造特征

从构造特征知识图谱（图 11a）可见，华南稀有金属矿床受印支运动和燕山运动形成的断裂控制。图谱中断裂构造主要方向集中在 NE、NNE 与 NW 向（图 11b），这恰好和印支期和燕山期在华南构造展布的方向一致。从知识图谱中可以看到，和成矿相关的构造除了断裂之外，还和褶皱、背斜、向斜相关，这进一步说明了和稀有金属成矿相关的动力学背景除了拉张环境，挤压造山可能也有一定的影响。



## Diagram of Common Fault Trends and Ore Body Morphology)

### (4) 华南稀有金属矿床蚀变和矿物组合

图谱结果显示，华南稀有金属矿床的蚀变作用普遍发育，类型复杂且空间分带明显。整体上，蚀变系统以钠长石化、绢云母化、云英岩化、硅化、萤石化、绿泥石化和黄玉化为重（图 12）。其中，钠长石化和云英岩化通常出现在成矿岩体的边部或上盘位置，是岩浆-流体相互作用最直接的产物；硅化与黄玉化常与高温富氟流体活动有关，指示成矿后期的强烈热液交代；而绢云母化与萤石化则代表中低温阶段的流体再活化与冷却沉淀。这种由高温钠长石化-云英岩化到低温绢云母化-萤石化的演化序列，反映了华南稀有金属成矿体系经历了明显的多阶段热液演化过程。

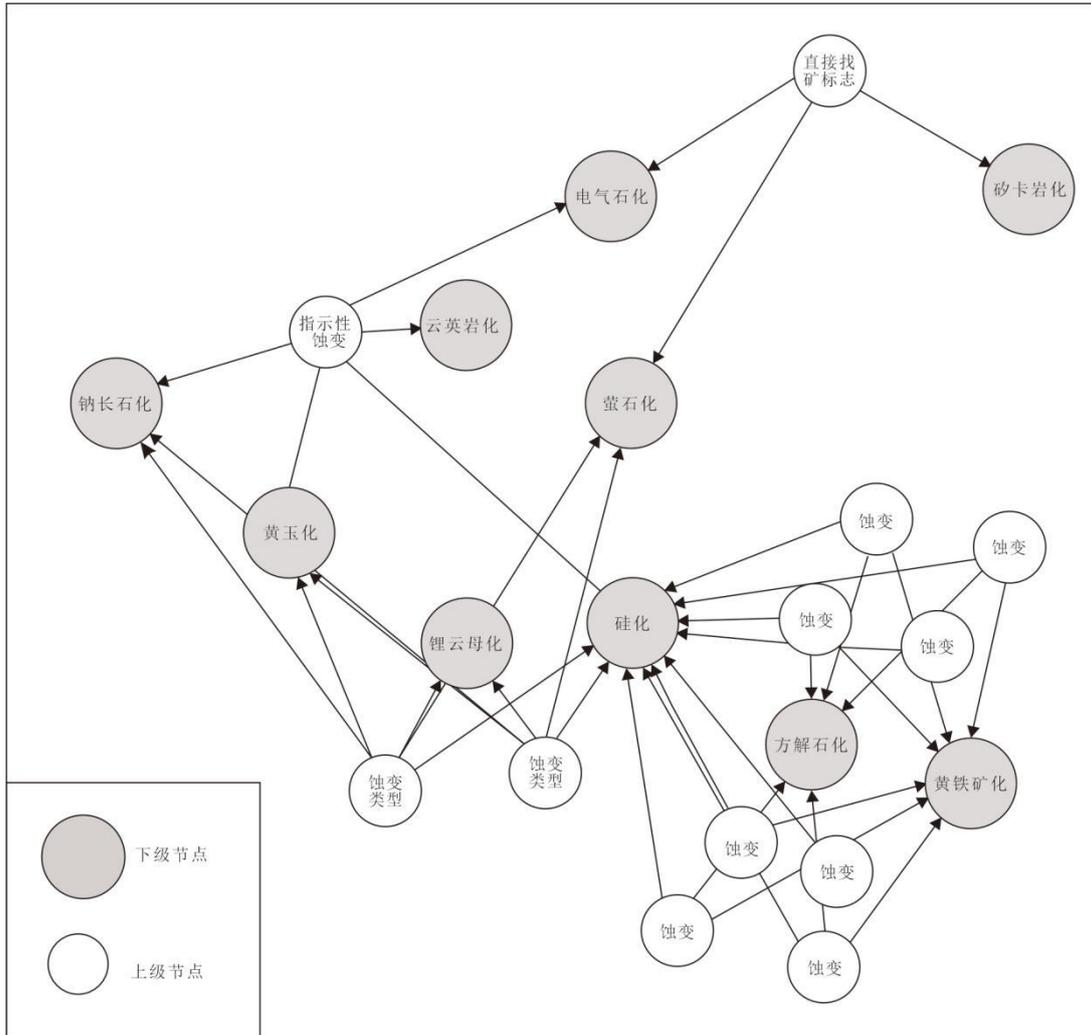


图 12. 华南稀有金属矿床（点）矿化类型知识图谱

Figure 12. Knowledge Graph of Mineralization Types in Rare Metal Deposits and Occurrences in South China

在矿物组合方面,图谱揭示出稀有金属矿床普遍具有多金属共生与矿物序列分带特征(图 13)。Li、Be、Nb、Ta 等元素与锂辉石、绿柱石、铌铁矿、钽铁矿、锡石、白钨矿等有关。图谱中具有岩浆从演化早期到晚期不同阶段出现的矿物,但从关系线的连接数量来看,最多的分别为长石、石英、萤石、绿柱石、黄玉、白钨矿、黑钨矿、锡石、锂辉石、锂云母、铌钽铁矿等。这种特征说明华南稀有金属矿床的形成经历了从花岗质岩浆结晶到挥发分释放、再到流体交代富集连续演化过程,蚀变作用不仅反映了成矿热液的性质和温度条件,也直接控制了稀有金属矿物的形成与富集位置。综合分析,华南稀有金属矿床的蚀变与矿物组合关系揭示了成矿体系的双阶段特征:早期岩浆分异控制金属供给与初始富集,晚期富氟热液交代可能决定了矿物种类与富集强度,两者共同构成了华南地区稀有金属成矿作用的关键地球化学机制。

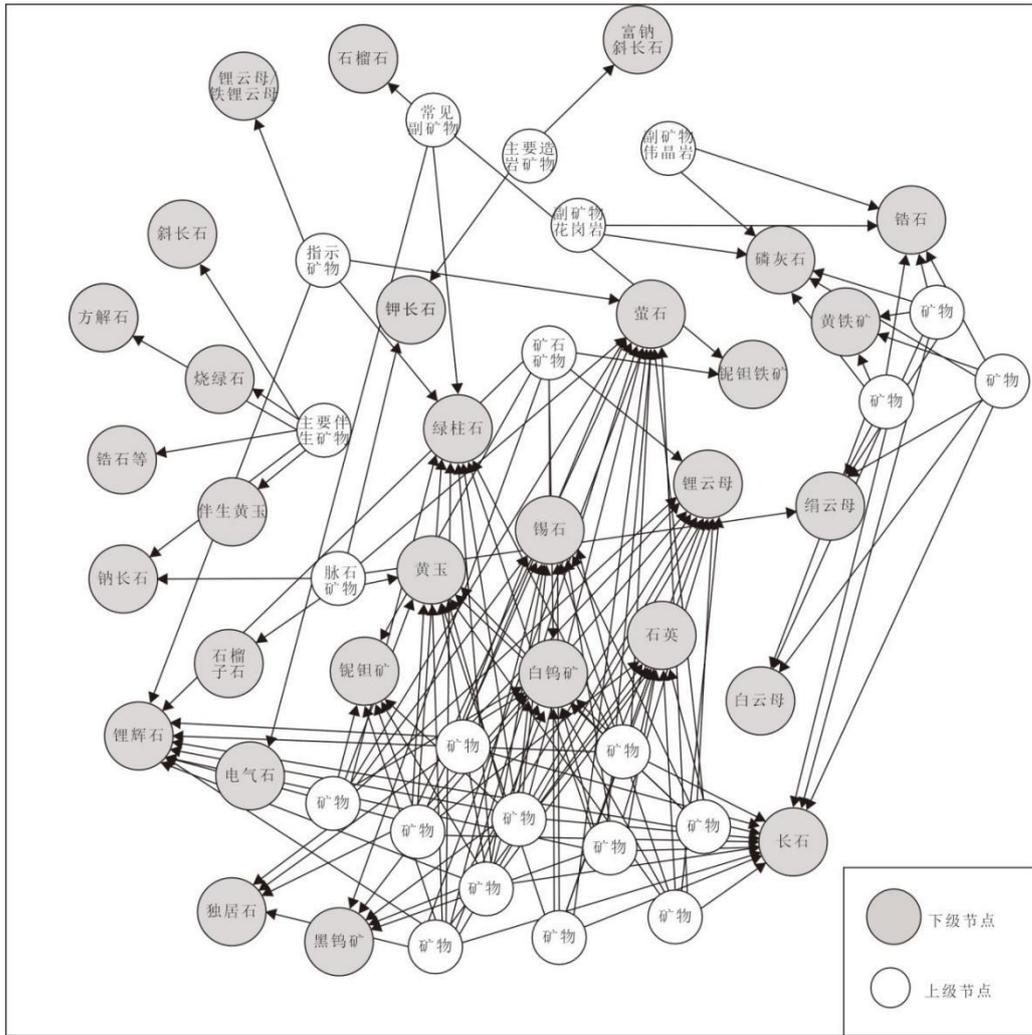


图 13. 华南稀有金属矿床（点）矿物类型知识图谱

Figure 13. Knowledge Graph of Mineral Types in Rare Metal Deposits and Occurrences in South China

### 3 结论

(1) 本研究基于 Deepseek R1-32B 大语言模型与提示词工程，构建了兼顾成矿机制与找矿标志的统一矿床知识本体，实现了从非结构化地质文献中自动化抽取实体与关系。结果表明，大语言模型能够有效识别复杂地质语义并生成符合地质逻辑的知识结构，为垂直领域知识图谱构建提供了可复制的技术路径。

(2) 华南稀有金属矿床的成矿作用与印支期、燕山期的岩浆活动密切相关，主要受高分异花岗岩及其富挥发分流体控制。成矿体系表现出明显的多阶段岩浆

-热液演化特征，稀有金属元素（Li-Be-Nb-Ta-W-Sn）在演化后期的流体阶段进一步富集，形成多金属共生与空间分带现象。

（3）知识图谱的结果揭示了稀有金属矿床在地球化学异常、构造控制及蚀变分带方面的内在联系，为区域找矿预测与成矿规律研究提供了新的认知方式和智能化支撑，展示了大语言模型在地学知识发现中的创新潜力。

## References

- Cheng,Q.M.,2021.What are Mathematical Geosciences and its frontiers? *Earth Science Frontiers*,28(03):006-025(in Chinese with English abstract).
- Cracknell,M.J., Reading,A.M.,2014.Geological mapping using remote sensing data: A comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information. *Computers & Geosciences*, 63: 22-33. doi:10.1016/j.cageo.2013.10.008
- Deng,Y.,Song,S.,Fan,J.,et al.,2024.Paleontology knowledge graph for data-driven discovery. *Journal of Earth Science*, 35(3): 1024-1034. doi:10.1007/s12583-023-1943-9
- Devlin,J.,Chang,M.-W.,Lee,K.,et al.,2019.BERT:Pre-training of deep bidirectional transformers for language understanding. *Proceedings of NAACL-HLT 2019*, 4171-4186.
- Dong,J.,Qiu,Q.,Xie,Z.,et al.,2023.Understanding table content for mineral exploration reports using deep learning and natural language processing. *Ore Geology Reviews*, 156: 105383. doi:10.1016/j.oregeorev.2023.105383
- Grover,A.,Leskovec,J.,2016.node2vec:Scalable feature learning for networks. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'16)*. ACM, New York, pp. 855-864. doi:10.1145/2939672.2939754
- Guan,N.,Song,D.,Liao,L.,2019.Knowledge graph embedding with concepts. *Knowledge-Based Systems*, 164: 38-44. doi:10.1016/j.knosys.2018.10.008
- Hou,Z.D.,Zhao,Z.,Liu,Z.J.,et al.,2023.Metallogenic regularity and prospecting direction of granite related Li-Be-Nb-Ta deposits in the Nanling region, South

- China. *Acta Petrologica Sinica*,39(7):1950-1972 (in Chinese with English abstract).
- Lahitani,A.R.,Permanasari,A.E.,Setiawan, N.A.,2016.Cosine similarity to determine similarity measure: Study case in online essay assessment. In: Proceedings of the 4th International Conference on Cyber and IT Service Management. pp. 1-6. doi:10.1109/CITSM.2016.7577578
- Lawley,C.J.M.,Gadd,M.G.,Parsa,M.,et al.,2023.Applications of natural language processing to geoscience text data and prospectivity modeling. *Natural Resources Research*, 32(4): 1503-1527. doi:10.1007/s11053-023-10216-1
- Li,L.,Zhou,R.,Huang,D.,2009.Two-phase biomedical named entity recognition using CRFs. *Computational Biology and Chemistry*, 33(4): 334-338. doi:10.1016/j.compbiolchem.2009.07.004
- Lima,L.A.,Görnitz,N.,Varella,L.E.,et al.,2017.Porosity estimation by semi-supervised learning with sparsely available labeled samples. *Computers & Geosciences*, 106: 33-48. doi:10.1016/j.cageo.2017.05.004
- Li,X.F.,Wei,X.L.,Zhu,Y.T.,et al., 2021.Rare metal deposits in South China:Types,characteristics,distribution and tectonic setting. *Acta Petrologica Sinica*,37(12):3591-3614(in Chinese with English abstract).
- Ma,X.,Hovy,E.,2016.End-to-end sequence labeling via bi-directional LSTM-CNNs-CRF. arXiv preprint, arXiv:1603.01354.
- Ma,X.,Wang,H.,Lehmann,B.,et al.,2024.Control of magmatic halogen composition and redox state on the zonation of metal mineralization across active continental margins: Perspectives from the world-class South China metallogenic province. *Chemical Geology*, 669: 122363. doi:10.1016/j.chemgeo.2024.122363
- Maynez,J.,Narayan,S.,Bohnet,B.,et al.,2020.On faithfulness and factuality in abstractive summarization. *Proceedings of ACL 2020*, 1906-1919.
- Ma,X.H.,Yan,J.Y.,Wang,H.H.,et al.,2025.Decoupling of tungsten and tin mineralization governed by magmatic fluorine. *Science China Earth Sciences*, 68(8): 2474–2491 (in Chinese).
- Ou,Q.,Liu,J.,Zi,F.,et al.,2025.Late Mesozoic Wangxiang composite granitic pluton,

- South China Block: Implications to magma emplacement and evolution from geochemical proxies. *Journal of Earth Science*, 36(2): 485-507. doi:10.1007/s12583-022-1760-8
- Qin,J.H.,Wang,D.H.,Wang,Y.,et al.,2023.Metallogeni law and exploration prospect of the middle part of Nanling metallogenic belt, south China block. *Geology in China*,1-26( in Chinese with English abstract).
- Song,S.,Mao,J.,Romer,R.L.,et al.,2024.Hosts of Sn in reduced deep-seated W skarn systems: A case study on the world-class scheelite skarn deposit, Zhuxi, South China. *Mineralium Deposita*, 59(7): 1433-1454. doi:10.1007/s00126-024-01271-6
- Wang,C.,Ma,X.,Chen,J.,et al.,2018.Information extraction and knowledge graph construction from geoscience literature. *Computers & Geosciences*, 112: 112-120. doi:10.1016/j.cageo.2017.12.007
- Wu,F.Y.,Guo,C.L.,Hu,F.Y.,et al.,2023.Petrogenesis of the highly fractionated granites andtheir mineralizations in Nanling Range, South China. *Acta Petrologica Sinica*,39(1):1-36(in Chinese with English abstract).
- Wu,Q.H.,Zhou,H.X.,Liu,B.,et al.,2023.Spatio-temporal distribution of granite-related rare metal deposits and W-Sn deposits in South China and their genetic relationship. *Bulletin of Geological Science and Technology*,42(1):78-88(in Chinese with English abstract).
- Xiao,F.,Chen,Z.,Chen,J.,et al.,2016.A batch sliding window method for local singularity mapping and its application for geochemical anomaly identification. *Computers & Geosciences*, 90: 189-201. doi:10.1016/j.cageo.2015.11.001
- Xiao,W.,Liu,C.,Tan,K.,et al.,2025.Zircon U–Pb–Hf and trace element signatures reveal the petrogenesis of the Jiuyishan granitic complex, South China: Implications for W–Sn and rare metal mineralization. *Journal of Earth Science*, 36(3): 1069-1089. doi:10.1007/s12583-023-1842-2
- Yan,Q.,Xue,L.,Li,Y.,et al.,2023.Mineral prospectivity mapping integrated with geological map knowledge graph and geochemical data: A case study of gold deposits at Raofeng area, Shaanxi Province. *Ore Geology Reviews*, 161: 105651.

doi:10.1016/j.oregeorev.2023.105651

Yang,J.,Wu,J.,Zhou,M.,et al.,2025.Mantle contributions to global tungsten recycling and mineralization. *Communications Earth & Environment*, 6(1): 510.  
doi:10.1038/s43247-025-02471-2

Zhang,Z.,Yang,Z.,Jian,F.,et al.,2025.Interpretability-enhanced mineral prospectivity models: A synergistic approach using large language models, knowledge graphs, and machine learning. *Mathematical Geosciences*.  
doi:10.1007/s11004-025-10231-3

Zhang,R.Z.,Zuo,R.G.,2025.Comparison of the geochemical characteristics of highly fractionated granites in South China and the Himalaya. *Acta Petrologica Sinica*,41(4):1424-1441(in Chinese with English abstract).

Zhang,Z.J.,Cheng,Q.M.,Yang,J.,et al.,2021.Machine learning for mineral prospectivity: A case study of iron-polymetallic mineral prospectivity in southwestern Fujian. *Earth Science Frontiers*,28(3):221-235(in Chinese with English abstract).

Zhao,M.,Zhang,Z.,Yang,J.,et al.,2025.Knowledge graph construction and knowledge discovery for porphyry copper deposits. *Ore Geology Reviews*, 186: 106875.  
doi:10.1016/j.oregeorev.2025.106875

Zhao,Y.,He,J.W.,Zhu,S.C.,et al.,2024. Security of Large Language Models:Current Status and Challenges. *Computer Science*,51(01): 68-71(in Chinese with English abstract).

Zhao,Z.,Chen,Y.C.,Wang,D.H.,et al.,2022.Transformation of Mesozoicdynamic systems and superposition of metallogenic series of W-Sn-Li-Be-Nb-Ta-REE mineral deposits in South China. *Acta Petrologica Sinica*,38(02):301-322(in Chinese with English abstract).

Zhou,Y.Z.,Zuo,R.G.,Liu,G.,et al.,2021.The great-leap-forward development of mathematical geoscience during 2010-2019: Big data and artificial intelligence algorithm are changing mathematical geoscience. *Bulletin of Mineralogy, Petrology and Geochemistry*, 28(3): 1-16(in Chinese with English abstract).

Zhu,Y.S.,2006.Basic theories of mineral prediction: A theoretical system extending

regional metallogeny to mineral exploration. *Acta Geologica Sinica*, 84(10): 1518-1527(in Chinese).

Zhu,Z.S.,Zhu,L.,1998.Prediction theory and method system of deposits. *Journal of Chengdu University of Technology*, S1: 6-12(in Chinese).

### 中文参考文献

- 成秋明. 2021. 什么是数学地球科学及其前沿领域? *地学前缘*, 28(03): 6-25.
- 侯占德, 赵正, 柳振江, 等. 2023. 南岭花岗岩区锂铍铌钽成矿规律与找矿方向. *岩石学报*, 39(7): 1950-1972.
- 李晓峰, 韦星林, 朱艺婷, 等. 2021. 华南稀有金属矿床: 类型、特点、时空分布与背景. *岩石学报*, 37(12): 3591-3614.
- 马星华, 闫金禹, 王宏晖, 等. 2025. 岩浆中的氟对钨锡成矿解耦的控制. *中国科学:地球科学*, 55(08): 2583-2602.
- 秦锦华, 王登红, 王岩, 等. 2023. 华南板块南岭成矿带中段成矿规律与找矿前景. *中国地质*, 1-26.
- 吴福元, 郭春丽, 胡方泂, 等. 2023. 南岭高分异花岗岩成岩与成矿. *岩石学报*, 39(1): 1-36.
- 吴玺虹, 周厚祥, 刘飏, 等. 2023. 华南与花岗岩有关的稀有金属矿床和钨锡矿床的时空分布规律及其成因联系. *地质科技通报*, 42(1): 78-88.
- 张睿泽, 左仁广. 2025. 华南和喜马拉雅高分异花岗岩地球化学特征对比分析. *岩石学报*, 41(4): 1424-1441.
- 张振杰, 成秋明, 杨玠, 等. 2021. 机器学习与成矿预测:以闽西南铁多金属矿预测为例. *地学前缘*, 28(03): 221-235.
- 赵月, 何锦雯, 朱申辰, 等. 2024. 大语言模型安全现状与挑战. *计算机科学*, 51(01): 68-71.
- 赵正, 陈毓川, 王登红, 等. 2022. 华南中生代动力体制转换与钨锡铍铌钽稀土矿床成矿系列的叠加演化. *岩石学报*, 38(02): 301-322.
- 周永章, 左仁广, 刘刚, 等. 2021. 数学地球科学跨越发展的十年: 大数据、人工智能算法正在改变地质学. *矿物岩石地球化学通报*, 40(03): 556-573+777.
- 朱裕生. 2006. 矿产预测理论——区域成矿学向矿产勘查延伸的理论体系. *地质*

学报, 84(10): 1518-1527.

朱章森, 朱磊. 1998. 矿床定量预测理论与方法体系. 成都理工学院学报, S1:  
6-12.