# 摘要

随着地质调查与研究工作的持续深入，地质领域积累了大量非结构化文本资料，如地质调查报告、矿产评估文献与学术论文等，蕴含着丰富的地质实体与实体间复杂多元关系信息。传统基于规则或二元三元组的关系抽取方法难以全面捕捉这些信息，特别是在表达复杂语义结构、地层部位限定、岩性描述等细粒度关系时，存在信息丢失、语义不完整等问题。

针对这一挑战，本文提出一种基于大语言模型的地质实体多元关系抽取方法，并构建了首个面向地质领域的多元关系标注数据集。该数据集涵盖 23 类实体类型、10 种主关系与 12 种限定键，来源于 4 篇地质调查报告，总计包含 2311 组多元关系事实，具有实体类型丰富、语义关系复杂、长尾关系显著等特点，能够为地质领域多元关系建模提供基础支持。

在此基础上，本文引入 Qwen2.5-7B-Instruct 大语言模型，结合结构化指令模板设计与语义相似样本检索机制，构建领域增强型的指令微调框架，实现端到端的联合实体识别与多元关系抽取。为进一步提升结果准确性与一致性，提出了模型自检的验证机制。实验结果表明，该方法在构建的数据集上取得了良好的性能，多元关系抽取任务的 F1 值达到 0.7474，较通用领域基线模型 Text2NKG 提升了 8.35%，验证了所提方法在地质语境下的有效性与适配性。

本研究为非结构化地质文本中复杂关系的高质量抽取提供了一种新路径，也为后续开展地质信息智能分析与辅助决策奠定了数据与模型基础。



**关键词**：地质实体；多元关系抽取；大语言模型；指令微调


# 1 引言

地质科学是研究地球组成、结构、演化过程及其相互关系的重要学科，其核心内容通常依赖于大量关于岩石、地层、构造等实体及其间语义联系的文本描述。这些信息广泛分布在地质调查报告、矿产资源勘查档案、科研论文等非结构化文本中，随着地质调查工作的不断推进，此类文本数据的规模持续增长[1]。然而，由于非结构化文本内容冗长、表达形式多样，传统依赖人工整理的方法效率低、易遗漏，已难以满足当前对高效、智能化地质信息处理的实际需求[2]。

关系抽取（Relation Extraction, RE）作为自然语言处理（Natural Language Processing, NLP）中的一项基础任务，致力于从文本中识别实体之间的语义关系[3]。当前主流方法多聚焦于提取二元关系，即以“头实体-关系-尾实体”的三元

组形式表达事实[4]。然而，在地质文本中，语义结构复杂、实体嵌套频繁，仅使用简单的二元关系难以覆盖实际语境中的丰富含义[5]。许多语义事实不仅包含多个核心实体，还依赖于描述性信息（如部位、年代、倾角、品位等）来准确界定关系内容。例如，“某岩层上部富含灰绿色细粒砂岩且品位为 5%”这样的表述，若仅抽取“岩层-含有-砂岩”的三元组，将丢失关键语义。研究表明[6]，现实文本中超过三分之一的关系事实涉及三元以上的交互关系，需通过多元关系建模加以表达。

多元关系抽取（N-ary Relation Extraction）正是为解决此类问题而提出的[6]。该方法以主三元组为基础，结合若干限定键值对（qualifier key-value pairs）以补充、限定和细化语义内容，从而实现对真实语境中复杂关系的准确建模。该类方法在生物医学[7]、金融[8]、法律[9]等领域已有广泛应用，取得了良好效果。但在地质领域，由于语料资源稀缺、术语高度专业化以及结构复杂性强等问题，相关研究仍处于起步阶段[5]。一方面，尚缺乏高质量的地质多元关系标注数据集作为训练与评估基础；另一方面，已有通用模型难以迁移至地质语境，模型输出易出现信息缺失或结构不一致的问题。

近年来，大语言模型（Large Language Models, LLMs）在多种自然语言任务中展现出卓越的表现，特别是在复杂语义建模、少样本学习等方面具备显著优势。如 GPT[10]、T5[11]、Qwen[12]等模型，通过指令微调（Instruction Fine-tuning）[13]与结构化提示设计（Prompt Engineering）[14]，已广泛用于文本生成、实体识别、关系抽取等任务[15]。然而，将 LLMs 应用于地质多元关系抽取任务仍存在困难[16]：一是模型可能因领域知识不足产生幻觉输出（hallucination），生成不准确或冗余的关系事实；二是复杂结构的输出控制难度较高，缺乏对实体角色、限定键约束等细节的精准把握。

基于此挑战，本文聚焦于地质领域非结构化文本中的复杂关系抽取任务，提出一种融合领域知识、结构化指令设计与大语言模型优化的多元关系抽取方法。首先，本文构建了首个系统性地质多元关系标注数据集，覆盖 23 类实体类型、10 种关系与 12 种限定键，从 4 篇地质调查报告中精细标注并整理了 1099 个样本，形成 2311 组高质量多元关系事实，为模型训练与评估提供了坚实基础。其次，设计了面向大语言模型 Qwen2.5-7B-Instruct 的结构化指令模板，结合语义相似示例检索策略，构建了领域增强型指令微调框架，有效引导模型输出符合格式规范、语义完整的多元关系结构。为提升结果一致性与模型可控性，本文进一步提出了验证机制，通过模型自检的方式识别并修正常见抽取误差，增强系统鲁棒性。实验结果表明，所提出方法在地质领域多元关系抽取任务中 F1 值达到 0.7474，较通用模型基线提升超过 8%，在实体识别、限定键抽取、关系结构完整性方面

均表现出良好的适配性。

总体而言，本文研究不仅系统提出并验证了一种适用于地质文本语义抽取的技术框架，更通过数据构建与模型设计填补了地质多元关系抽取领域的研究空白。本研究为后续地质信息处理、语义分析与辅助系统开发奠定了关键基础，展现出良好的实用前景与学术价值。

# 2 相关工作

关系抽取作为自然语言处理领域的核心任务，旨在从非结构化文本中识别实体之间的语义关联，构建结构化三元组形式的知识表示。随着深度学习技术的发展，研究者逐渐从早期的基于规则与模板的方法转向依赖神经网络模型的端到端抽取架构。其中，BERT（Bidirectional Encoder Representations from Transformers）模型的提出[17]为关系抽取任务带来了显著突破。

## 2.1 BERT 在关系抽取中的应用

在关系抽取任务中，BERT 的双向编码特性使其能够同时感知头实体与尾实体所处的上下文环境，因此被广泛应用于句级和篇章级的关系抽取中。例如，Zhan J 等人[18]将 BERT 与 CNN 结合用于句子级别的二元关系抽取任务，显著提高了模型对语法模式的感知能力；Chen B 等人[19]进一步引入多通道注意力机制，增强模型对实体之间隐含语义的捕捉能力。

尽管 BERT 在关系抽取任务中已展现出优越性能，但其原生设计仍主要面向二元关系建模，难以直接扩展至 n 元关系抽取场景。地质文本中存在大量结构复杂、实体关系多样的描述，仅依赖头尾实体和单一关系的三元组表达，往往无法完整呈现知识语义。例如，“某岩层中上部含灰绿色细粒砂岩”这类描述需要引入“位置”、“颜色”、“粒度”等附加信息方能实现精确表达。此类语义需求超出 BERT 模型设计范式，限制其在地质领域多元关系抽取中的表现。

## 2.2 大语言模型与多元关系抽取

随着模型规模与训练数据的急剧扩展，大语言模型如 GPT-3、T5、Qwen 等逐渐成为推动自然语言理解和生成任务的重要引擎。与传统 Encoder-only 架构的 BERT 不同，LLMs 多采用 Decoder-only 或 Encoder-Decoder 混合架构，具备强大的语言建模能力和跨任务迁移泛化能力，能够支持复杂的生成式输出场景。近年来，LLMs 在信息抽取、文档理解、数据增强等领域中表现出良好的迁移性

能，逐渐被引入至关系抽取任务中。

特别是随着指令微调与提示工程技术的成熟，研究者开始尝试使用 LLMs 以零样本（Zero-Shot）或少样本（Few-Shot）方式执行复杂结构信息抽取任务。Zhu W 等人[20]采用 ChatGPT 对金融类文档中的句级二元关系进行抽取，展示了模型在通用领域下的零样本迁移能力；Ren P 等人[21]则尝试将多元关系抽取任务建模为机器阅读理解任务，通过结构化提示设计引导 LLM 输出事件型多元结构。然而上述工作多数聚焦于通用语料场景，缺乏对地质领域文本中专业术语、多实体结构及上下位层级等特征的适配与优化。

当前，LLMs 在地质实体多元关系抽取中的应用仍处于起步阶段。存在的主要问题包括：（1）语义泛化不足：现有模型训练语料以通用领域为主，面对地质类术语或地层结构等专业知识常产生“幻觉”现象，输出虚构或不连贯内容；（2）结构输出不一致：生成式 LLM 虽可通过自然语言生成三元组或多元组，但输出格式不稳定，影响结构化信息的解析；（3）关系粒度表达受限：多数模型仅建模头尾实体及关系，缺乏对限定性附加信息（如“部位”、“倾角”、“品位”等）的表达机制，无法满足地质知识图谱对细粒度语义的需求。

# 3 方法

## 3.1 任务定义

地质实体多元关系抽取任务的目标是从地质领域的非结构化文本中，提取包含两个以上实体的复杂关系，并以结构化的形式表示。本文采用超关系（Hyperrelation）模式定义多元关系抽取任务[22]，其核心是通过主三元组（头实体-关系-尾实体）结合附加的限定键值对，实现对复杂关系的细粒度描述。任务的数学定义如下：

给定一个地质文本句子集合 D，预定义的关系类别集合 R，以及限定键类别集合 Q，任务是从每个句子 $s \in D$ 中抽取格式化的多元关系，结构表示为：

$$G = \left(h, r, t, (q_1, v_1), (q_2, v_2), \ldots, (q_n, v_n)\right) \tag{1}$$

其中：

(1) $h$：头实体（Head Entity），如“遮拉组（J2z）”；

(2) $r \in R$：关系类型，如“含有”；

(3) $t$：尾实体（Tail Entity），如“绢云板岩”；

(4) $(q_i, v_i)$：限定键值对，其中 $q_i \in Q$ 为限定键（如“地层部位”），$v_i$ 为限定值（如“中上部”），限定键值对数量 $n \geq 1$。

(5) 主三元组 $(h,r,t)$ 描述核心关系，限定键值对 $(q\_i,v\_i)$ 提供附加语义信息，用于补充和细化主三元组的语义。例如，对于句子“遮拉组（J2z）地层的中上部层位，含有绢云板岩”，抽取的多元关系为：

$$\left(\text{遮拉组,含有,绢云板岩,}\left(\text{地层部位,中上部}\right)\right) \tag{2}$$

## 3.2 数据集构建

### 3.2.1 数据源选择

本文数据集来源于中国地质调查局地质云网站中的地质调查报告，选取了四篇具有代表性的报告作为数据源。这四篇报告分别为《雄村铜矿总报告》（197 个样本）[23]、《西藏 1:25 万尼玛区幅、热布喀幅区域地质调查报告》（443 个样本）[24]、《琼果幅、曲德贡幅(1:5 万)地质调查报告》（381 个样本）[25]以及《洛阳市非金属矿产资源》（78 个样本）[26]。这些报告涵盖了多种地质实体（如岩石、地层、矿物、构造等）及其复杂关系，文本内容专业性强、语义丰富，能够充分反映地质领域非结构化文本的特点。样本总数为 1099 个，划分为训练集（880 个样本，约占 80%）和测试集（219 个样本，约占 20%），以支持模型的训练与评估。

### 3.2.2 关系类型和限定键

为实现细粒度的地质实体多元关系抽取，本文定义了 23 种实体类型、10 种关系类型以及 12 种限定键。

实体类型包括岩石、地质时代、地层、构造等，覆盖地质文本中的核心概念。10 种关系类型分别为：<*含有*>、<*位于*>、<*出露于*>、<*厚度*>、<*所属年代*>、<*沉积构造形态*>、<*整合接触*>、<*不整合接触*>、<*倾向*>、<*假整合*>。每种关系均有明确的语义定义，例如“含有”描述某岩石地层中包含特定岩石或矿物，“出露于”表示地质实体在某地区的暴露情况。

12 种限定键用于补充主三元组（头实体-关系-尾实体）的语义信息，包括<*位于方位*>、<*出露面积*>、<*出露方位*>、<*地层部位*>、<*时代前后*>、<*含量和品位*>、<*岩石性质描述为*>、<*岩石构造形态为*>、<*岩石晶型为*>、<*倾角*>、<*断层的某个部分*>、<*地层上下位置关系*>。例如，“地层部位”限定键可描述岩石出现在地层的中上部或下部，“含量和品位”限定键用于说明矿物的含量占比。这些限定键的引入弥补了二元关系抽取在细粒度语义表达上的不足，使抽取结果更符合地质领域的实际需求。

### 3.2.3 数据预处理

数据预处理包括文本清洗、格式标准化和分句处理三个步骤。首先，对原始

地质调查报告进行文本清洗，移除无关内容（如页眉、页脚、表格标题等），保留与地质实体和关系相关的核心文本。其次，将文本统一转换为 UTF-8 编码格式，并对标点符号、空格等进行规范化处理，确保数据格式一致性。最后，采用基于句号和换行符的分句策略，将长篇报告拆分为独立句子，生成 1099 个样本句子，便于后续的实体和关系标注。分句过程中，保留上下文信息以避免语义割裂，例如，确保同一段落中涉及相同地质实体的句子保持关联性。

### 3.2.4 数据标注

数据标注使用开源标注工具 Label-studio[27]，从每条样本中标注出所有实体（包括实体类型和文本边界）、主三元组关系以及限定键值对。标注规范基于预定义的实体和关系，确保一致性。例如，对于句子“遮拉组（J2z）地层的中上部层位，岩石类型有绢云板岩”，标注结果为：主三元组{遮拉组，含有，绢云板岩}，限定键值对{地层部位，中上部}。

"测区位于青藏高原中部偏南，地处著名的雅鲁藏布江结合带与特提斯喜马拉雅带衔接部位。"
位于 测区 地名和行政区划 0-2 青藏高原 地名和行政区划 4-8 位于方位 中部偏南 具体方位 8-12

"测区地层主要出露有前震旦系、古生界、三叠系、侏罗系、白垩系等地层，它们位于青藏高原中部偏南，相差较大。"
位于 前震旦系 地层、组、段 9-13 青藏高原 地名和行政区划 37-41 位于方位 中部偏南 地层部位描述 41-45
位于 古生界 地层、组、段 14-17 青藏高原 地名和行政区划 37-41 位于方位 中部偏南 地层部位描述 41-45
位于 三叠系 地层、组、段 18-21 青藏高原 地名和行政区划 37-41 位于方位 中部偏南 地层部位描述 41-45
位于 侏罗系 地层、组、段 22-25 青藏高原 地名和行政区划 37-41 位于方位 中部偏南 地层部位描述 41-45
位于 白垩系 地层、组、段 26-29 青藏高原 地名和行政区划 37-41 位于方位 中部偏南 地层部位描述 41-45

"分布于喜马拉雅地层区（康马-隆子分区）的羊卓雍地层小区的三叠系-白垩系地层出露面积约181km2，占整个测区面积的20%。"
出露于 羊卓雍地层小区 地层区 20-27 三叠系-白垩系地层 地层、组、段 28-37 出露面积 约181km2 面积描述 41-48

"亚堆扎拉岩组横向上以浅灰色-云斜长片麻岩、黑云斜长片麻岩、二云石英片岩为主夹斜长变粒岩。"
含有 亚堆扎拉岩组 地层、组、段 0-6 二云斜长片麻岩 岩石和矿物 13-18 岩性为 浅灰色 岩性描述词 19-13
含有 亚堆扎拉岩组 地层、组、段 0-6 斜长片麻岩 岩石和矿物 21-26 岩性为 黑云 岩性描述词 19-21

图 1 数据集示例

最终生成如图 1 所示的包含 2311 组多元关系的高质量数据集，其中每个样本被空行分隔，被“”包裹的内容是样本，而在样本下面几行中存储的则是多元关系事实，数据的组织形式为：[关系][头实体][头实体类型][头实体位置][尾实体][尾实体类型][尾实体位置][限定键][限定值][限定值类型][限定值位置]。

其中三元关系占 77.15%，四元关系占 21.9%，五元及以上关系占不到 1%，如图 2 所示。此外，数据集包含 2864 个限定键值对，反映了地质文本中关系的复杂性和多样性。

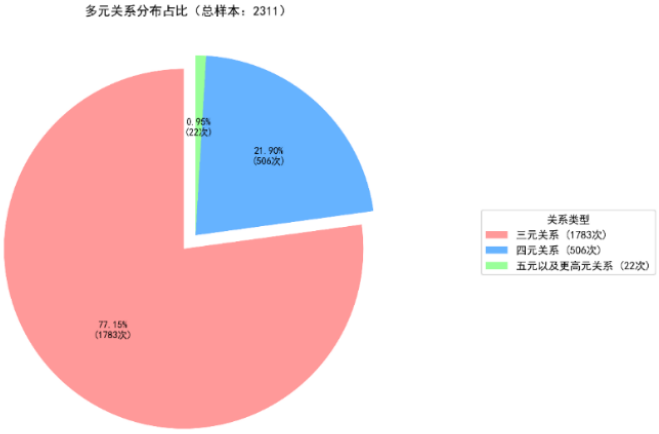


图 2 多元关系分布情况

## 3.3 基于大语言模型的地质实体多元关系抽取框架

本文提出了一种基于大语言模型的地质实体多元关系抽取框架，以 Qwen2.5-7B-Instruct 模型[12]为基座，通过指令设计、模型微调和双重验证机制实现从非结构化地质文本中提取结构化多元关系事实。框架的整体流程如图 3，以下详细介绍框架的三个核心组成部分。

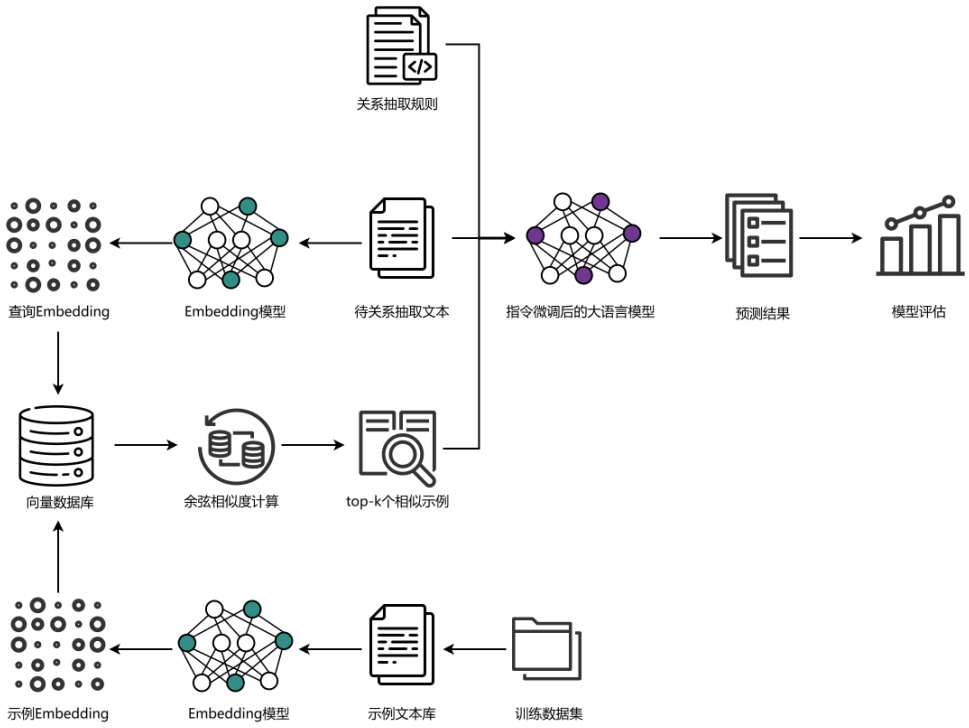


图 3 框架整体流程图

### 3.3.1 指令设计

指令设计是引导大语言模型准确理解和执行多元关系抽取任务的关键。本文通过精心设计的结构化提示词，结合地质领域规则和语义相似样本检索策略，提升模型对复杂关系的抽取能力。指令设计包括抽取规则介绍、检索示例样本和结构化提示词三个子模块，完整的指令示例如图 4 所示。

已知的关系类型为：<含有>、<位于>、<出露于>、<厚度>、<所属年代>、<沉积构造形态>、<整合接触>、<不整合接触>、<倾向>、<假整合>。

已知的限定词键为：<位于方位>、<出露面积>、<出露方位>、<地层部位>、<时代前后>、<含量和品位>、<岩石性质描述为>、<岩石构造形态为>、<岩石晶型为>、<倾角>、<断层的某个部分>、<地层上下位置关系>

针对于关系倾向对应的尾实体必须是一个角度；针对于限定键倾角对应的限定值必须是一个角度；针对于限定键品位对应的限定值必须是一个百分数；当头尾实体之间既有位于关系又有出露于关系时只保留出露于。

请找出文本中的三元组和限定词的键值对，JSON模板为：{"三元组": (头实体, 关系类型, 尾实体), "限定词表": [{限定词键: 限定词值}]}，其中头实体、尾实体、限定词值必须来源于文本，关系类型和限定词键必须来源于已知内容。

文本中的三元组可能不止一种，示例：
"input": "测区地层主要出露有前震旦系、古生界、三叠系、侏罗系、白垩系等地层，它们位于青藏高原中部偏南，相差较大。"
"output": [{"三元组": {"前震旦系", "位于", "青藏高原"}, "限定词表": [{"位于方位": "中部偏南"}]}, {"三元组": {"古生界", "位于", "青藏高原"}, "限定词表": [{"位于方位": "中部偏南"}]}, {"三元组": {"三叠系", "位于", "青藏高原"}, "限定词表": [{"位于方位": "中部偏南"}]}, {"三元组": {"侏罗系", "位于", "青藏高原"}, "限定词表": [{"位于方位": "中部偏南"}]}, {"三元组": {"白垩系", "位于", "青藏高原"}, "限定词表": [{"位于方位": "中部偏南"}]}]"

限定词的键值对也可能不止一条，示例：
"input": "白松岩段下部为中厚—厚层块状砂岩，向上以中—中薄层为主，板岩下部少，向上增多，其中见有少量黄铁矿，鲍玛序列总体不发育，局部发育A-E段，多为D-E段，仅发育重荷模沉积构造。"
"output": [{"三元组": {"白松岩段", "含有", "砂岩"}, "限定词表": [{"地层部位": "下部"}, {"岩石性质描述为": "中厚—厚层块状"}]}, {"三元组": {"白松岩段", "含有", "板岩"}, "限定词表": [{"地层部位": "下部"}]}]"

以下是类似的抽取案例：
案例1：
案例2：
案例3：

下面给出我的文本：

图 4 多元关系抽取指令

1. 抽取规则介绍

抽取规则通过关系类型库、限定键库、特殊实体约束和冲突消解规则四个组件明确任务要求。关系类型库和限定键库分别列出 10 种关系和 12 种限定键，确保模型理解每种关系的语义范围。

特殊实体约束组件针对地质文本中常见的歧义实体（如地层名称与行政区划名称重叠）设置规则，例如“*针对于关系倾向对应的尾实体必须是一个角度；针对于限定键倾角对应的限定值必须是一个角度；针对于限定键品位对应的限定值必须是一个百分数*”。

冲突消解规则组件解决多关系样本中的冲突问题，例如“*当头尾实体之间既有位于关系又有出露于关系时只保留出露于*”。这些规则的引入是为了降低模型的幻觉输出，提升抽取结果的准确性。

2. 检索示例样本

为增强模型对复杂地质关系的泛化能力，本文采用基于语义相似度的少样本学习策略，从训练集中检索与输入样本语义相近的示例。具体方法是使用预训练的嵌入向量模型计算输入样本与训练集样本的余弦相似度[28]，选取相似度最高的 3 个样本作为提示词中的示例。示例样本包含完整的多元关系标注（如主三元组和限定键值对），为模型提供明确的输入-输出映射关系。例如，对于输入句子“永珠组在措勤－南木林南部”，检索到的示例为“白垩纪花岗岩出露于格仁错构造带其南侧”，其标注结果为{白垩纪花岗岩，出露于，格仁错构造带，出露方位，南侧}。语义相似样本的引入有助于模型更好地理解地质文本的语义模式，减少对低频关系的误判。

3. 结构化提示词

结构化提示词由任务描述、输入数据、示例和输出格式四部分组成，旨在规范模型的输出。任务描述清晰定义多元关系抽取的目标，要求模型从输入句子中提取主三元组及限定键值对，并以 JSON 格式输出：“*请找出文本中的三元组和限定词的键值对，JSON 模板为：{“三元组”: (头实体, 关系类型, 尾实体), “限定词表”: [{限定词键: 限定词值}]}*”。输入数据为待抽取的原始句子，示例部分提供少样本学习的输入-输出对，输出格式指定为标准化的 JSON 结构。结构化输出格式便于后续程序解析，同时确保抽取结果的规范性和一致性。

### 3.3.2 模型微调

模型微调采用全量参数微调（Full Parameter Fine-tuning）[13]方法，基于 Qwen2.5-7B-Instruct 模型和 Llama-Factory 框架[29]，在地质实体多元关系抽取数据集上进行监督学习。微调过程以指令-输出对的形式输入训练数据，其中指令为结构化提示词，输出为标注的 JSON 格式多元关系事实。微调目标是最小化交叉熵损失函数，通过反向传播更新模型全部参数，使模型更好地适配地质领域的专业术语和复杂语义结构。

为提升微调效果，本文设计了领域增强型指令模板，将地质领域规则（如关系定义、限定键约束）嵌入指令中，引导模型聚焦于地质文本的语义特征。此外，通过动态调整训练样本的顺序和多样性，增强模型对长尾关系的泛化能力。微调后的模型能够端到端地完成联合实体和关系抽取任务，显著提升了在地质领域低资源场景下的抽取性能。

### 3.3.3 双重验证机制


已知的关系类型为：<含有>、<位于>、<出露于>、<厚度>、<所属年代>、<沉积构造形态>、<整合接触>、<不整合接触>、<倾向>、<假整合>。

已知的限定词键为：<位于方位>、<出露面积>、<出露方位>、<地层部位>、<时代前后>、<含量和品位>、<岩石性质描述为>、<岩石构造形态为>、<岩石晶型为>、<倾角>、<断层的某个部分>、<地层上下位置关系>。

重要规则说明：
1. 倾向关系的尾实体必须是角度（如 205°）；
2. 整合接触/不整合接触/假整合的实体必须是地层或岩石；
3. 同时存在位于和出露于关系时优先保留出露于；
4. 地层部位的限定值应为：顶部、上部、中部、下部、底部；
5. 倾角限定值必须是角度格式（如 ∠65°）；
6. 品位限定值必须是百分数（如 60%）；
7. 位于方位限定值只能是：东、南、西、北、东北、西北、东南、西南。

已提取的预测结果格式为：{三元组: (头实体, 关系类型, 尾实体), 限定词表: [{限定词键: 限定词值}]}。
实体和值必须来自文本，关系和键必须来自上述定义。

请执行以下验证：
1. 检查关系和限定键是否符合上述定义；
2. 确认所有实体均来自原文；
3. 检查实体是否符合要求；
4. 是否存在遗漏的多元关系。
如果多元关系整体抽取正确则说明原因；如果多元关系中的实体、关系、限定词键、限定词值出现错误则给出更好的结果。

判断结果类型有：<多元关系正确>、<多元关系错误>、<多元关系缺失>。
错误类型有：
1. 无错误：多元关系正确；
2. 头实体错误：头实体不符合关系定义要求或不属于原文；
3. 关系类型错误：在头尾实体正确的情况下关系类型判断错误；
4. 尾实体错误：尾实体不符合关系定义要求或不属于原文；
5. 限定词键错误：在限定词值判断正确的情况下限定词键判断错误；
6. 限定词值错误：限定词值不符合限定词键的要求或不属于原文。
一个多元关系当中的错误可能不止一种。

你输出的 JSON 模板为：{
"原始三元组": (头实体, 关系类型, 尾实体),
"原始限定词表": [{限定词键: 限定词值}],
"判断结果": [{错误类型: 原因}],
"正确三元组": (头实体, 关系类型, 尾实体),
"正确限定词表": [{限定词键: 限定词值}]
}

图 5 抽取结果验证指令

为确保抽取结果的准确性和可解释性，本文提出了一种双重验证机制，通过设计专门的验证提示词（如图 5 所示），要求模型对初次抽取结果进行二次判断。验证提示词包括关系类型库、限定键库、规则说明、数据格式说明、验证要求、判断结果库和格式化输出七个组件。双重验证中定义了六种类型，如表 1 所示。模型根据提示词重新分析抽取结果，输出判断结果、错误原因及修正建议，并以 JSON 格式返回。

表 1 双重验证评价方式定义

| 编号 | 含义 |
|---|---|
| 1 | 代表在双重验证中大模型认为样本预测的结果没有问题的，同时抽取结果符合标注。 |
| 2 | 代表在双重验证中大模型认为样本预测的结果有问题，但是给出的原因是错误的。 |
| 3 | 代表在双重验证中大模型认为样本预测的结果有问题给，出的原因是讲得通的，但是修改后的结果是不符合真实标注的。 |
| 4 | 代表代表在双重验证中大模型认为样本预测的结果有问题，并且修改之后的结果是符合真实标注的。 |
| 5 | 代表大模型没有正确的给出双重验证结果。 |
| 6 | 大模型没有正确的判断出错误样本。 |

双重验证机制通过模型自检提高了抽取结果的可解释性。验证过程中发现的错误类型（如幻觉输出、限定值偏差）为后续模型优化提供了重要依据。

# 4 实验和分析

## 4.1 评估指标

在地质实体多元关系抽取任务中，模型性能的评估是衡量其效果的关键环节。本研究采用准确率（Precision）、召回率（Recall）和 F1 值作为主要评估指标，以全面衡量模型在实体、关系、限定键、限定值及整体多元关系抽取上的表现。这些指标的定义如下：

准确率：表示模型预测为正且真实值为正的样本占所有预测为正样本的比例，公式为：

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3}$$

召回率：表示模型预测为正的样本占所有真实值为正样本的比例，公式为：

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{4}$$

F1 值：综合考虑准确率和召回率，反映模型的整体性能，公式为：

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5}$$

其中，TP（True Positive）表示真值为正且预测为正的样本数量，FP（False Positive）表示真值为负但预测为正的样本数量，FN（False Negative）表示真值为正但预测为负的样本数量。

在地质实体多元关系抽取任务中，实体、关系、限定键和限定值的抽取准确性直接影响整体多元关系的正确性。因此，本研究分别计算了上述各部分的准确率、召回率和 F1 值，以分析模型在不同子任务中的表现。此外，针对地质实体名称的复杂性（如长实体、嵌套实体），采用模糊匹配算法评估实体抽取效果，设置不同精确度阈值，以平衡严格匹配与语义正确性之间的关系。实验中发现，精确度为 0.7 时能够较好地平衡抽取效果与语义完整性，因此后续分析主要基于此阈值。

## 4.2 实验设置

### 4.2.1 基线模型

为评估本文提出的基于 Qwen2.5-7B-Instruct 的多元关系抽取模型性能，选用

Text2NKG 模型作为基线模型。Text2NKG 由 Luo H 等人[22]提出，该模型包含以下核心模块：

(1) 数据增强：通过调整实体顺序生成多种排列（如交换头尾实体或调整辅助实体顺序），提升模型对不同实体排列的鲁棒性。例如，对于多元关系{A, r, B, [q, C]}，可生成 6 种排列方式。

(2) 异序合并：将不同顺序的预测概率合并为统一顺序（如{A, B, C}），提高预测一致性。

(3) 输出合并：将主三元组相同的限定键值对合并为更高元的多元关系事实，实现动态元数抽取。

(4) 空标签权重：通过设置超参数平衡空标签与其他标签的学习，优化交叉熵损失函数。

为适配地质领域，Text2NKG 的 BERT 模块替换为 BERT-base-Chinese，训练参数包括学习率 2e-5、批大小 8、训练轮次 10、dropout 0.1、空标签权重 0.01。

#### 4.2.2 实验参数

实验硬件环境基于 8 张 NVIDIA Tesla V100 显卡构建的 GPU 服务器，单卡显存为 32GB。

在模型微调阶段，采用 Llama-Factory 提供的全量参数微调功能，优化器选择 AdamW，初始学习率设置为 5e-5，训练总轮数为 5 轮，采用线性学习率调度策略，warmup 比例为 0.03，梯度裁剪阈值设置为 1.0，最大输入 token 长度为 2048。训练过程中启用混合精度（fp16）加速以降低显存占用，并使用 deepspeed[30]实现多卡并行训练。

推理阶段采用 vLLM 框架[31]部署 Qwen2.5-7B-Instruct 微调后模型，启用 continuous batching 与 prefill 优化机制，显著降低推理延迟。生成参数方面，设置最大生成长度为 1024，temperature 为 0.7，top_p 为 0.95，top_k 为 40，repetition_penalty 为 1.0，以避免重复输出与生成幻觉。

### 4.3 实验结果

#### 4.3.1 主实验结果

实验在自建的地质实体多元关系数据集上进行，该数据集包含 1099 个样本，划分为训练集（880 个样本）和测试集（219 个样本），涵盖 23 种实体类型、10 种关系类型和 12 种限定键，总计 2311 组多元关系事实。本文对比了以下模型的性能：Text2NKG、Qwen2.5 和提示词结合相似样本的 Qwen2.5，实验结果如表 2 所示：

表 2 模型实验结果

| 方法 | 类型 | Precision | Recall | F1 |
|---|---|---|---|---|
| Text2NKG | 多元关系 | 0.6991 | 0.632 | 0.6639 |
| Qwen2.5 | 多元关系 | **0.7428** | **0.7521** | **0.7474** |
| | 实体 | 0.9239 | 0.9423 | 0.933 |
| | 关系 | 0.8724 | 0.8833 | 0.8778 |
| | 限定键 | 0.978 | 0.935 | 0.956 |
| | 限定值 | **0.9385** | **0.946** | **0.9422** |
| 提示词结合相似样本的 Qwen2.5 | 多元关系 | 0.7423 | 0.75 | 0.7461 |
| | 实体 | **0.9454** | **0.9463** | **0.9458** |
| | 关系 | **0.9052** | **0.9146** | **0.9098** |
| | 限定键 | **0.9787** | **0.937** | **0.9574** |
| | 限定值 | 0.9075 | 0.9093 | 0.9084 |

从表 2 可以看出，本文提出的基于大语言模型的地质实体多元关系抽取框架优于基线模型 Text2NKG，具体分析如下：

1. 基线模型表现：Text2NKG 在通用领域表现优异，但在地质领域适配性较差，即使经过预训练，F1 值仅为 0.6639。这表明 BERT-base-Chinese 在处理地质领域复杂语义和长尾关系时存在局限，数据增强策略无法完全适应地质文本的专业性。

2. Qwen2.5 模型表现：提示词结合相似样本的 Qwen2.5 在实体（F1=0.9458）、关系（F1=0.9098）和限定键（F1=0.9574）抽取上均优于其他模型，但在限定值抽取上略逊于 Qwen2.5（F1=0.9422）。这可能由于语义相似样本引入的扰动导致限定值预测范围不准确，但整体性能仍显著优于基线模型，F1 值提升约 8.35%。

### 4.3.2 双重验证结果

双重验证结果如图 6 所示：

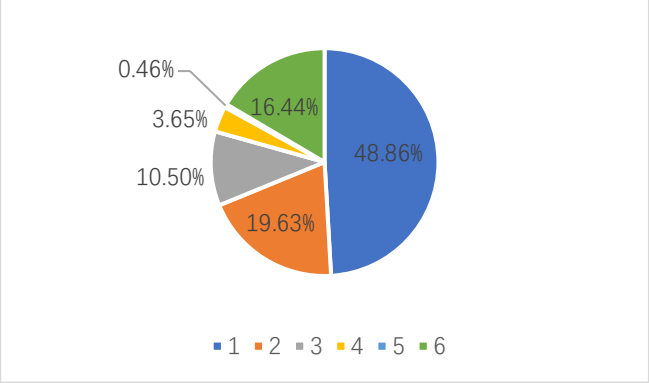


图 6 双重验证统计结果

1. 抽取正确，验证亦认为正确：占比 48.86%，这表明近一半的样本在首次抽取时就已达到正确结果，且通过模型自身的验证机制也得到了确认，说明所构建的指令框架在多数场景下具有较强的抽取能力与结构理解能力。

2. 抽取正确，验证误判为错误：占比 19.63%，该类别表明大模型在自检阶段对已有正确抽取结果的判断出现误差，主要源于验证指令中规则解释不够明确或错误类型定义模糊，导致模型对“是否为错误”缺乏清晰判断标准。

3. 抽取错误，验证判断为错误，但修正结果仍不正确：占比 10.5%，此类样本反映出模型虽然具备错误识别能力，但在修正建议方面仍显不足，未能生成与真实标注一致的结构。这一问题说明模型在遇到复杂错误时，缺乏对错误上下文的深入理解，验证模块需进一步引导模型生成结构完整、语义精确的修正结果。

4. 抽取错误，验证判断正确且修正正确：占比 3.65%，该类是验证机制理想目标的体现，即能够识别并修复抽取错误。但其占比较低，说明当前的自检机制尚处于初级阶段，模型对错误的定位与改正能力仍有限。

5. 验证阶段输出格式错误：占比 0.46%，虽然比例极低，但这表明少量样本在验证输出过程中存在格式异常，可能是由特殊字符、句法异常或模板泛化失效导致。

6. 抽取错误但验证未发现：占比 16.44%，该类是当前验证机制中最值得关注的短板，表明存在相当数量的抽取错误未被模型在验证阶段识别出来，显示验证机制的敏感性不足。这可能是由于验证所用指令与抽取指令结构过于接近，缺乏从另一个角度审视输入的能力。

## 4.4 案例分析

为更具体地揭示模型在实际抽取过程中的表现与差异，本文选取了三个具有代表性的样本进行深入案例分析，以评估各模型在面对不同地质文本结构与语义难点时的应对能力。

### 4.4.1 尾实体幻觉生成

如图 7 所示，三个模型在处理该句子时均出现了主三元组尾实体预测错误的情况，均误将尾实体识别为“石墨大理岩”，而该实体在原始文本中并不存在。此类错误属于典型的大语言模型“幻觉”问题，即模型在生成过程中凭借语言模式生成不真实或不相关的实体。在此案例中，即便引入了结构化指令与地质规则，模型仍未能有效抑制幻觉的产生。值得注意的是，融合语义相似样本的 Qwen2.5 模型在整个测试集中仅出现过两次类似幻觉，显示其在实体预测的稳定性方面具

有一定优势。

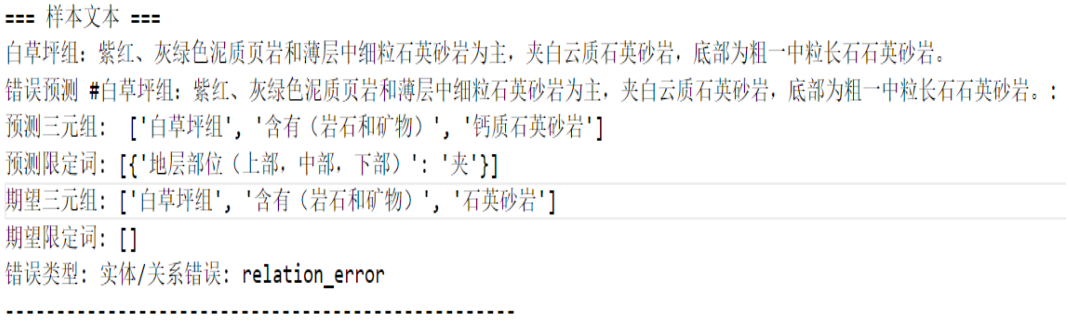


图 7 案例 1

### 4.4.2 尾实体识别的精确性提升

图 8 所示的案例进一步验证了语义相似示例对于模型性能的正向引导作用。在该句子中，仅有提示词结合相似样本的 Qwen2.5 方法成功识别了正确的尾实体。分析其原因，在于所引入的语义示例中包含结构相似、语义接近的句式与关系结构，使得模型在上下文理解与实体定位过程中获得了更明确的语义锚点，从而避免了错误抽取。这表明，语义检索增强机制不仅提升了模型的上下文对齐能力，也在一定程度上缓解了长尾关系与不常见实体带来的挑战。

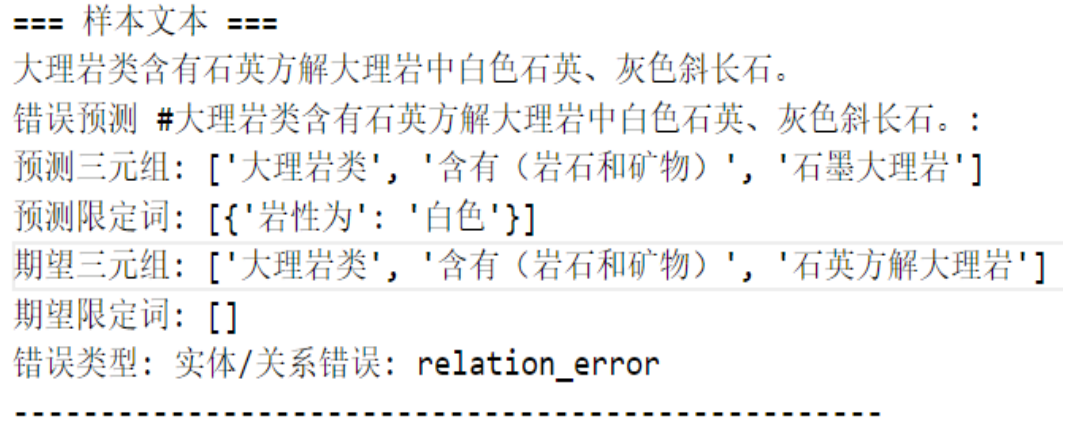


图 8 案例 2

### 4.4.3 复杂限定键匹配能力的差异

图 9 展示的案例表明，在包含多个限定键值对的多元关系样本中，基于大语言模型的方法能够成功抽取所有目标关系，而基线模型 Text2NKG 则出现了明显的错误。具体分析显示，Text2NKG 由于依赖 BERT 的双向注意机制，面对限定值与目标实体间距离较远的情况时，存在注意力分散与关系错配的问题。例如，某些限定键错误地关联到了邻近但无关的实体，导致关系条目出现结构性错误。而大语言模型在生成式架构下更倾向于依赖前文语义流，能够形成较为连贯的实体—限定键—值链条，表现出更优的结构保持能力。

在纳木错一带，昂杰组下部为灰绿色泥质粉砂岩、粉砂质泥岩夹凝灰质粉砂岩、灰岩，在林周一带为黑色板岩夹凝灰质砂岩，上部为灰岩夹钙质砂岩，厚约 170m。

关系条目 1

主体：昂杰组

关系类型：含有（岩石和矿物）

客体：凝灰质粉砂岩

地层部位：下部、上部

关系条目 2

主体：昂杰组

关系类型：含有（岩石和矿物）

客体：泥质粉砂岩

岩性：灰绿色

地层部位：下部

关系条目 3

主体：昂杰组

关系类型：含有（岩石和矿物）

客体：灰岩

地层部位：上部、下部

关系条目 4

主体：昂杰组

关系类型：含有（岩石和矿物）

客体：钙质砂岩

地层部位：上部

关系条目 5

主体：昂杰组

关系类型：含有（岩石和矿物）

客体：粉砂质泥岩

地层部位：下部

图 9 案例 3

综上所述，通过案例分析可以观察到以下结论：

1. 大语言模型在避免幻觉与处理复杂多元结构方面相较传统方法更具优势；

2. 引入语义相似样本作为提示示例能有效提升模型在细粒度语义区分上的敏感性；

3. 生成式结构有助于保持限定键值与目标实体之间的关联完整性。

上述发现进一步验证了本文方法在地质实体多元关系抽取任务中的实际可行性与鲁棒性。

# 5 结论

本文围绕地质领域非结构化文本中的复杂关系建模问题，系统提出了一种融合大语言模型、结构化指令设计与自建数据集的地质实体多元关系抽取框架。针对传统方法难以处理多实体、多限定信息的局限性，本文在方法论、数据资源与验证机制三个方面展开了深入研究，并取得了以下主要成果：

1. 首次构建地质多元关系抽取标注数据集。从四篇权威地质调查报告中精细筛选与加工，定义了 23 类实体、10 种关系与 12 类限定键，构建出 2311 组高质量的多元关系事实，填补了地质领域结构化语义资源的空白，并为模型训练与评估提供了坚实基础。

2. 提出基于 Qwen2.5-7B-Instruct 的大语言模型抽取方法。通过构建结构化指令模板、嵌入地质规则与语义示例检索机制，引导模型生成标准化、语义完整的多元关系输出。在此基础上，设计双重验证机制，引导模型对自身输出进行结构

与语义自检，有效提升了抽取结果的稳定性与可解释性。

3. 实验证明所提方法在多项指标上均优于基线模型。在自建测试集上，与 Text2NKG 模型相比，本文方法的 F1 值提升了约 8.35%。同时，在实体、关系、限定键与限定值四个子任务上均取得显著提升，展现出优越的语义建模能力与对地质专业术语的适配性。特别是在处理长尾关系、结构嵌套等复杂场景时，大模型展现出更强的泛化与容错能力。

综上，本文不仅验证了大语言模型在地质专业语义抽取任务中的有效性，也为复杂实体关系的知识图谱构建提供了可行路径。未来工作将围绕以下几个方向进一步拓展：一是扩充数据集规模与覆盖范围，适配更多地质文献语境；二是融合图像、图表等多模态信息，增强模型对地质场景的综合理解能力；三是探索原型系统的部署与应用，服务于地质调查、资源评估与辅助决策等实际需求。

# 参考文献


[1] 马凯.地质大数据表示与关联关键技术研究[D].中国地质大学,2019.

[2] 孙琛皓,庄子浩,焦守龙.基于注意力机制与 PCNN 的地质关系抽取方法[J].计算机与数字工程,2024,52(06):1795-1801.

[3] Zhao X, Deng Y, Yang M, et al. A Comprehensive Survey on Relation E xtraction: Recent Advances and New Frontiers[J]. arXiv preprint arXiv:230 6.02051, 2023.

[4] Wu T, You X, Xian X, et al. Towards deep understanding of graph convo lutional networks for relation extraction[J]. Data & Knowledge Engineering, 2024, 149: 102265.

[5] Tian M, Ma K, Wu Q, et al. Joint extraction of entity relations from geol ogical reports based on a novel relation graph convolutional network[J]. C omputers & Geosciences, 2024, 187: 105571.

[6] Noy N , Rector A , Hayes P ,et al.Defining N-ary Relations on the Sema ntic Web[J]. 2006.

[7] Whitton J, Hunter A. Automated tabulation of clinical trial results: A joint entity and relation extraction approach with transformer-based language re presentations[J]. Artificial intelligence in medicine, 2023, 144: 102661.

[8] Bose P, Srinivasan S, Sleeman IV W C, et al. A survey on recent named entity recognition and relationship extraction techniques on clinical texts[J]. Applied Sciences, 2021, 11(18): 8319.

[9] Sasidharan A K, Rahulnath R. Structured approach for relation extraction i n legal documents[C]//2023 4th IEEE Global Conference for Advancement in Technology (GCAT). IEEE, 2023: 1-6.

[10] Brown T, Mann B, Ryder N, et al. Language models are few-shot learners [J]. Advances in neural information processing systems, 2020, 33: 1877-19 01.

[11] Raffel C, Shazeer N, Roberts A, et al. Exploring the limits of transfer lea rning with a unified text-to-text transformer[J]. Journal of machine learning research, 2020, 21(140): 1-67.

[12] Bai J, Bai S, Chu Y, et al. Qwen technical report[J]. arXiv preprint arXiv: 2309.16609, 2023.

[13] Chung H W, Hou L, Longpre S, et al. Scaling instruction-finetuned langua ge models[J]. Journal of Machine Learning Research, 2024, 25(70): 1-53.

[14] Reynolds L, McDonell K. Prompt programming for large language models: Beyond the few-shot paradigm[C]//Extended abstracts of the 2021 CHI co nference on human factors in computing systems. 2021: 1-7.

[15] Zhu W, Wang X, Chen X, et al. Refining ChatGPT for Document-Level Relation Extraction: A Multi-dimensional Prompting Approach[C]//Internatio nal Conference on Intelligent Computing. Singapore: Springer Nature Singa pore, 2024: 190-201.

[16] Wang X, Zhao W, Zhu X, et al. Can ChatGPT Solve Relation Extraction? An Extensive Assessment via Design Choice Exploration[C]//CCF Internat ional Conference on Natural Language Processing and Chinese Computing. Singapore: Springer Nature Singapore, 2024: 346-358.

[17] Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirection al transformers for language understanding[C]//Proceedings of the 2019 con ference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short paper s). 2019: 4171-4186.

[18] Zhan J, Zhao H. Span model for open information extraction on accurate corpus[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2020, 34(05): 9523-9530.

[19] Chen B, Yuan G. Domain knowledge-driven relation extraction methods[J]. Journal of Data Science and Intelligent Systems, 2024.

[20] Zhu W, Wang X, Chen X, et al. Refining ChatGPT for Document-Level Relation Extraction: A Multi-dimensional Prompting Approach[C]//International Conference on Intelligent Computing. Singapore: Springer Nature Singapore, 2024: 190-201.

[21] Ren P, Xu T, Qu J, et al. Tuning n-ary relation extraction as machine reading comprehension[J]. Neurocomputing, 2023, 562: 126893.

[22] Luo H, Yang Y, Yao T, et al. Text2nkg: Fine-grained n-ary relation extraction for n-ary relational knowledge graph construction[J]. Advances in Neural Information Processing Systems, 2024, 37: 27417-27439.

[23] 康丛轩. 西藏自治区雄村斑岩型铜(金)矿矿床特征及资源潜力评价[D]. 四川:成都理工大学,2011.

[24] 刘成社. 西藏 1:25 万尼玛区幅、热布喀幅区域地质调查报告[J]. 河南国土资源,2003(6):35.

[25] 曾庆高,毛国政,王保弟,等. 琼果幅、曲德贡幅(1:5 万)地质调查新成果及主要进展[J]. 地质通报,2004,23(5):475-478.

[26] 石毅,洛阳市非金属矿产资源研究.河南省,洛阳市国土资源局,2012-11-24.

[27] Tkachenko M, Malyuk M, Holmanyuk A, et al. Label Studio: Data labeling software[EB/OL]. (2020-2025) [2025-06-26]. https://github.com/HumanSignal/label-studio.

[28] Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks

[29] Zheng Y, Zhang R, Zhang J, et al. Llamafactory: Unified efficient fine-tuning of 100+ language models[J]. arXiv preprint arXiv:2403.13372, 2024.

[30] Jeff Rasley, Samyam Rajbhandari, Olatunji Ruwase, and Yuxiong He. (2020) DeepSpeed: System Optimizations Enable Training Deep Learning Models with Over 100 Billion Parameters. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '20, Tutorial).

[31] Kwon W, Li Z, Zhuang S, et al. Efficient memory management for large language model serving with pagedattention[C]//Proceedings of the 29th Symposium on Operating Systems Principles. 2023: 611-626.