

# 基于 Spring Boot 和 Vue 的地质信息 智能抽取与可视化系统设计与实现

邱芹军<sup>1)</sup>, 吕云峰<sup>1)</sup>, 吴麒瑞<sup>1)</sup>, 田苗<sup>1)</sup>, 孙江兵<sup>2,3)</sup>, 陶留锋<sup>1)</sup>, 张燕<sup>3)</sup>

1) 中国地质大学(武汉)计算机学院, 武汉, 430074;

2) 河海大学地球科学与工程学院, 南京, 211100;

3) 江苏省地质资料馆, 南京, 210012

**内容提要:**长期以来地质调查及信息化建设积累了海量的多源异构地质数据,而信息处理仍依赖人工录入和分散分析,导致数据利用效率低、挖掘不够、决策滞后。为解决已有通用抽取工具抽取效率低等问题,笔者等设计了一套智能信息抽取与可视化系统。该系统采用前后端分离架构,引入多种大语言模型作为信息抽取底座,结合领域知识通过自定义模板,从多模态复杂地质数据中智能化抽取高置信结构化信息,并通过表征学习与相似度量实现地质数据多特征关联,形成地质知识图谱,最终利用 ECharts 进行多维度交互式可视化。该系统实现了从文本到知识的转化,可自动生成知识图谱以揭示实体联系,为地质文本的智能化处理与深度利用提供了有效的解决方案,通过融合前沿人工智能与可视化技术,显著提升了地质知识的可用性与发现效率,为构建智能化地质知识服务平台奠定了坚实的实践基础。

**关键词:**地质知识图谱;大语言模型;软件工具;多模态地质数据

随着地球科学的不断发展,地质矿产勘探、地质灾害等领域积累了数量庞大的文本资料,包括地质志、区域勘探报告、学术论文。这些文本资料中蕴含着丰富的地质知识与时空信息,是研究区域地质演化规律和指导实际工作的关键依据。该部分信息大多以文本模态存储,对于依赖人工阅读和整理的传统处理方式上效率和准确性都较为低下(周永章等,2021; Zhang Xueying et al., 2022; 邱芹军等,2024; 周圣荃等,2025)。随着地球科学研究迈向深部、深地、深空,这种方式难以支撑对大规模文本进行快速、系统的分析。近年来,信息抽取和知识图谱等技术发展为地质文本的深入挖掘利用提供了新的方法。信息抽取技术旨在从非结构化文本中自动识别出命名实体、关系和事件等关键信息(刘文聪等,2021; 董家慧子等,2023; Dagdelen John et al., 2024; 周永章等,2024)。而知识图谱作为一种语义网络,能够将分散的信息有机整合,揭示实体之间的深层联系,形成可供推理与分析的知识网络(Qiu Qinjun

et al., 2023a; 王远超等,2024; 张超等,2025; 张泽琛等,2025)。这为地质资料的系统化、结构化利用开辟了新的路径,并已在多个领域展现出巨大潜力(Qiu Qinjun et al., 2023b; 邱芹军等,2023, 2024; Zhang Jinbin et al., 2025)。

已有信息抽取工具如 Knowledge Graph Studio 等大多数面向单一任务如面向实体、关系属性等抽取,缺少针对地质领域专业数据抽取的智能化工具,该类工具仅仅面向单一功能,分散的系统平台造成难以整合同一地区多源异构数据(如地质图件、勘探报告等),造成信息孤岛现象;其次,传统方法缺乏智能化分析能力,对复杂地质规律的识别仍依赖经验判断,准确性受限(王堃屹等,2024; 王汉雨等,2024)。同时,数据格式标准化不足,使得跨平台共享和长期保存面临挑战,最终影响地质信息的应用价值和社会化服务能力(Huang Zongcai et al., 2025)。针对上述问题,笔者等设计并实现面向地质领域的智能信息抽取与可视化系统,该系统以深

注:本文为国家自然科学基金资助项目(编号:42301492),江苏省自然资源厅地质数据智能应用技术创新中心开放基金项目(编号:GDATIC-XM-202502)和中国地质调查局地质调查项目(编号:DD20240029)的成果。

收稿日期:2025-10-10; 改回日期:2026-02-08; 网络首发:2026-02-15; 责任编辑:刘志强。Doi: 10.16509/j.georeview.2026.02.081

作者简介:邱芹军,男,1988年生,博士,副研究员,主要从事地质大数据挖掘与知识图谱研究;Email: qiuqinjun@cug.edu.cn。通讯作者:孙江兵,男,1986年生,硕士,高级工程师,主要从事地质资料综合利用、地质数据挖掘与应用等研究;Email: 230209020002@hhu.edu.cn。



度学习模型为基础,创新性地引入大语言模型(Large Language Model, LLM)来处理领域地质复杂信息,力图实现从地质文本规范化表示、结构化表达及语义化关联。系统集成信息抽取、知识图谱构建、多模态检索与可视化分析等功能模块,旨在为地质研究人员提供一个高效、直观辅助工具,提高地质信息整理和知识发现的效率。本系统通过前后端分离架构实现3大提升:①在系统架构方面,采用微服务动态连接数据库的机制,确保用户始终获取数据库中最新的地学要素核心数据;②在可扩展方面,系统功能模块支撑其他领域或细分地质领域多模态信息智能抽取与关联融合;③在算法方面,系统配置了支撑深度学习模型和大语言模型等各类接口,可快速接入系统实现多模态地质信息快速智能抽取。

## 1 系统总体设计

### 1.1 系统总体架构

本系统设计遵循当下主流的Web应用开发模式,整体上采用前后端分离的架构(图1)。该架构能够明确区分用户交互与业务逻辑:前端侧重界面呈现和可视化,后端则承担数据处理、模型调用和持久化管理。二者通过统一API接口实现通信,从而保证开发过程并行性与系统运行稳定性。对于地质信息处理这类对数据处理能力与可视化展示均有较高要求应用而言,该分层架构能够在保持模块独立性同时,实现灵活扩展和后续维护。

#### 1.1.1 技术选型与框架

在具体技术选择上,系统结合了成熟度、易用性

及在科研场景适用性进行权衡:

(1)后端采用Spring Boot构建。该框架不仅能快速搭建服务,还集成内置Tomcat容器,便于系统独立运行和部署。系统利用Spring MVC来实现RESTful API接口,从而为前端提供统一的数据服务。在数据库交互方面,选用MyBatis-Plus,在保持MyBatis灵活性同时,简化了常见的CRUD操作,能显著降低开发成本。特别值得一提的是,核心的信息抽取功能通过HTTP方式调用本地部署的LLMs模型,这使得后端业务逻辑与模型推理解耦,便于后续替换或升级不同模型。

(2)前端选择Vue 3框架。其引入的Composition API在组织复杂逻辑和复用组件方面更为灵活,同时结合Vue Router实现单页面应用(SPA)的平滑导航。界面组件基于Element Plus,既提高了开发效率,也保证了界面的交互一致性。在可视化部分,系统引入了ECharts库,用以绘制知识图谱和多特征关联图。这类基于Canvas的图表库在大规模节点渲染时依然能保持良好性能,适合展示复杂的地质实体关系。

(3)数据库选用MySQL作为核心存储。考虑到系统需要存放用户上传的文本、抽取结果及知识图谱中的节点和关系,MySQL的稳定性与社区支持使其成为可靠的选择。

这一技术选型的组合,既保证了开发过程中的可操作性,也为系统在科研场景中的长远使用提供了技术保障。

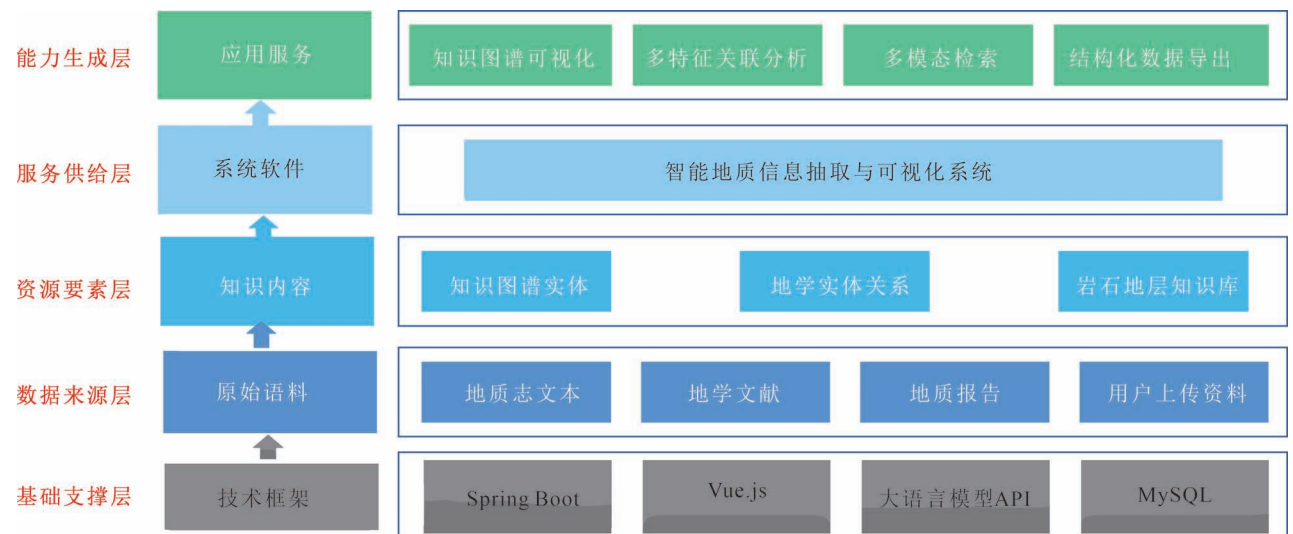


图1 系统总体架构示意图

Fig. 1 System architecture overview diagram

### 1.1.2 前后端分离架构设计

在架构设计上,前端和后端之间通过 RESTful API 进行通信,数据采用 JSON 格式交换(图 2)。这种设计不仅便于模块间的独立开发与测试,也使系统具备更强的灵活性:无论是前端界面的改版,还是后端的模型更新,都能相对独立地推进。

在部署方面,系统前端静态资源可直接放置在常见的 Web 服务器(如 Nginx 或 Apache)上进行托管,以保证访问的便捷性和稳定性。后端服务则以 Spring Boot 打包后的独立应用形式运行,便于在不同环境下快速部署和调试。总体而言,该系统的架构既吸收 Web 工程化的成熟经验,又结合地质信息抽取与可视化的特殊需求,形成兼顾灵活性与科研实用性的技术框架。

## 1.2 系统创新点

### 1.2.1 灵活智能化地质信息抽取

系统在地质领域引入大语言模型驱动的“模板化配置”抽取方法。用户可自定义抽取字段并传递至模型执行,支持多模型并行调用与一致性检查,从而在复杂语义条件下实现高精度、鲁棒性的结构化信息抽取,解决了非结构化文本难以利用的问题。

### 1.2.2 多源异构数据的语义检索与关联融合

设计“填图对象主题索引树”机制,将地质资料中的主题词、段落位置与空间位置信息建立关联,实现基于语义与空间位置的高置信度检索。同时,系统支持文本、图件、表格等多源异构数据的整合与联动,突破传统单一查询方式的限制,提升了地质知识的系统化表达与探索能力。

### 1.2.3 动态知识图谱与多模态可视化交互

提出“人机协同—动态生成”的知识库构建模

式,将抽取结果经人工校验后入库,并基于 ECharts 实现力导向布局的交互式知识图谱展示。结合多模态检索模块,系统实现了文本—图像的联动展示与双向验证,使用户能够直观探索复杂实体关系和地质图件,提高了知识的可视化表达和认知效率。

## 1.3 数据与知识资源

### 1.3.1 地质文本数据来源与特征

本系统所处理的数据主要来自用户上传的地质报告(如区域地质调查报告、矿产资源调查报告、地质志、期刊文献等)。作为一种综合性区域地质资料,区域地质调查报告详细记载了研究区的地层、岩石、构造、岩浆活动及矿产分布等信息,具有显著的专业性和完整性。与一般的自然语言文本相比,地质报告在以下几个方面表现出独特特征:

(1)术语高度集中。文本中包含大量地质学专有名词,例如地层单元(如“天井坪岩组”)、岩石类型(如变粒岩、石英岩)、矿物成分以及地质年代等。这种专业词汇的密集出现,对地质信息智能抽取与对齐融合等提出了更高要求。

(2)关系表述复杂。地质报告往往通过长句或嵌套句式来描述地层的分布范围、形成过程以及结构特征。这些信息不仅涉及空间位置和时间序列,还隐含了实体间的多重关系,直接增加了自动化抽取的难度。

(3)不同地区和编纂年代的地质报告在风格和表述习惯上存在差异。即便遵循统一的编写规范,仍会在术语使用、句式结构等方面有所不同。这种差异性使得模型在面对跨区域文本时容易出现泛化不足的问题,也为信息抽取与知识整合带来了挑战。

### 1.3.2 领域知识库构建

系统的核心目标之一,是将非结构化文本转化为可存储和利用的结构化知识。为此,本研究提出了一种“人机协同、动态生成”的知识库构建模式。

该系统以用户上传的区域地质调查报告为数据源,调用大语言模型自动识别并抽取出预先设定的关键信息字段,例如地层名称、地层符号、分布区域和岩石组合特征等。抽取结果首先以表格(Excel)形式返回用户,用户可对其中的条目进行核验和修改,从而保证数据质量。

在用户确认后,这些结构化结果被导入数据库,形成知识库的基本单元。数据库按照“实体—关系”模式进行存储:实体表保存地层、岩石等基本对象,关系表则记录它们之间的联系。当前端需要进行可视化展示时,系统会从数据库中检索相关数据,

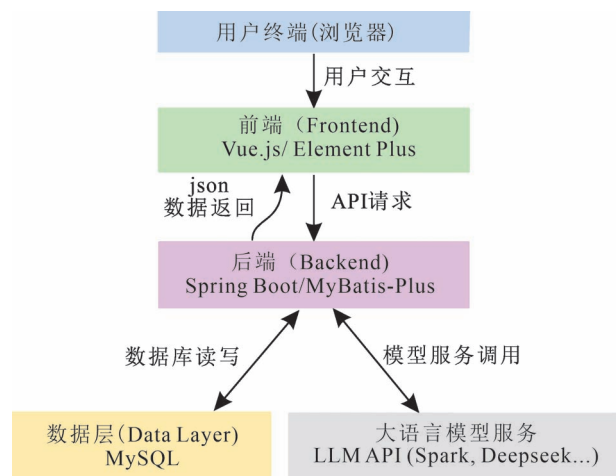


图 2 系统前后端分离架构示意图

Fig. 2 Frontend-backend decoupled architecture diagram

并通过 ECharts 渲染为知识图谱,帮助用户直观地理解不同地质要素之间的关联。

这种动态构建方式兼顾了自动化与专业性,一方面利用模型实现了高效抽取,另一方面保留了人工校正环节,以避免错误传播。最终形成的岩石地层知识库,不是静态的固定资源,而是可随研究目标与数据来源不断更新和扩展的知识体系,为后续的检索、可视化和分析奠定了基础。

## 2 系统功能

### 2.1 基于大模型的地质信息抽取模块

本系统以岩石地层为案例进行智能抽取。该模块创新性地引入大语言模型作为信息抽取的引擎,通过灵活的“模板化”配置,引导模型从非结构化的区域地质调查报告文本中,精准智能抽取出用户所需信息。

#### 2.1.1 地质文本数据预处理

本系统面向地质专业用户,简化了传统自然语言处理中复杂的预处理流程。系统的数据输入接口设计得十分直观,用户可以直接将本地的 .txt 格式地质报告文件导入,或将文本内容复制粘贴到指定的输入区域。系统在接收到文本后,主要进行基础的格式清理,如去除多余的空行和不可见字符,以确保输入到大语言模型中的文本序列是干净、规整的。整个预处理过程对用户透明,用户无需关心分词、词性标注等中间步骤,可以将精力完全集中在定义抽取任务和分析抽取结果上。

#### 2.1.2 信息抽取流程与实现

本系统的信息抽取实现了一个高度灵活的、由

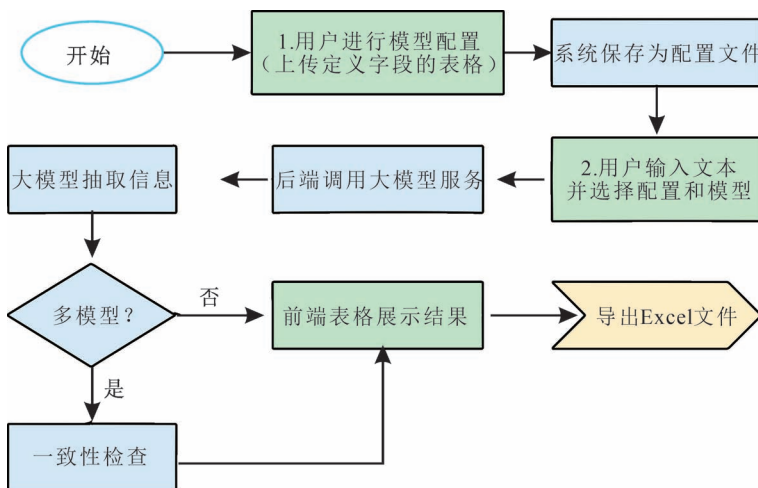


图3 岩石地层信息抽取流程图

Fig. 3 Lithostratigraphic information extraction workflow diagram

用户主导的人机协同流程。具体流程如图3所示。

(1)用户需要进行模型配置。这是指导大模型进行信息抽取的关键步骤。用户通过界面上上传一个预先定义好的表格文件,该表格的表头即定义了需要抽取的全部信息字段,例如:地层名称、地层符号、层型剖面经度、岩性特征、创建时间等。系统将此表头信息解析并保存为一个可复用的“配置文件”。用户可以创建多个不同的配置文件以适应不同的抽取任务,并对历史配置进行管理,包括修改、删除和导出。

(2)进入抽取执行阶段。用户在文本输入区导入待处理的地质志原文,然后在配置区选择一个先前创建好的配置文件,并勾选一个或多个希望使用的大语言模型(系统集成了 Llama、Spark Max、Spark4.0 Ultra、DeepSeek 等多种模型)。点击“开始”按钮后,前端将文本、配置文件标识和所选模型列表发送至后端。后端服务根据配置文件生成结构化的指令(Prompt),连同原始文本一并发送给指定的大语言模型服务。

(3)结果呈现与校验。大模型根据指令对文本进行分析,并返回 JSON 格式的结构化抽取结果。后端接收并解析后,将数据返回给前端,前端以表格的形式清晰地展示抽取结果,表格的列与用户配置文件中的字段一一对应。特别地,当用户选择多个模型时,系统会并行执行抽取,并在结果展示区提供“一致性检查”功能。该功能会自动比对不同模型抽取结果,并将存在差异的字段高亮提示出来,极大便利用户对结果进行交叉验证和择优。用户确认无误后,可将最终结构化数据一键导出为 Excel 文件,

完成从非结构化文本到结构化知识的转化。抽取实现截图如图4所示。

### 2.2 知识图谱可视化与检索模块

将抽取出的结构化地质信息转化为语义强关联的知识网络,是本系统的重要目标。本模块负责将存储在数据库中的实体和关系数据进行可视化呈现,并提供便捷的交互式探索与检索功能。

#### 2.2.1 可视化展示设计

用户将在信息抽取模块获得的 Excel 文件导入本模块后,系统后端会解析文件内容,将数据存入 MySQL 数据库的实体表与关系表中。前端随即向后端请求图谱数据,并调用 ECharts 的图表渲染引擎进行绘制。系统采用力引导布局 (Force -



图 4 实体关系抽取效果图  
Fig. 4 Entity relationship extraction effect diagram

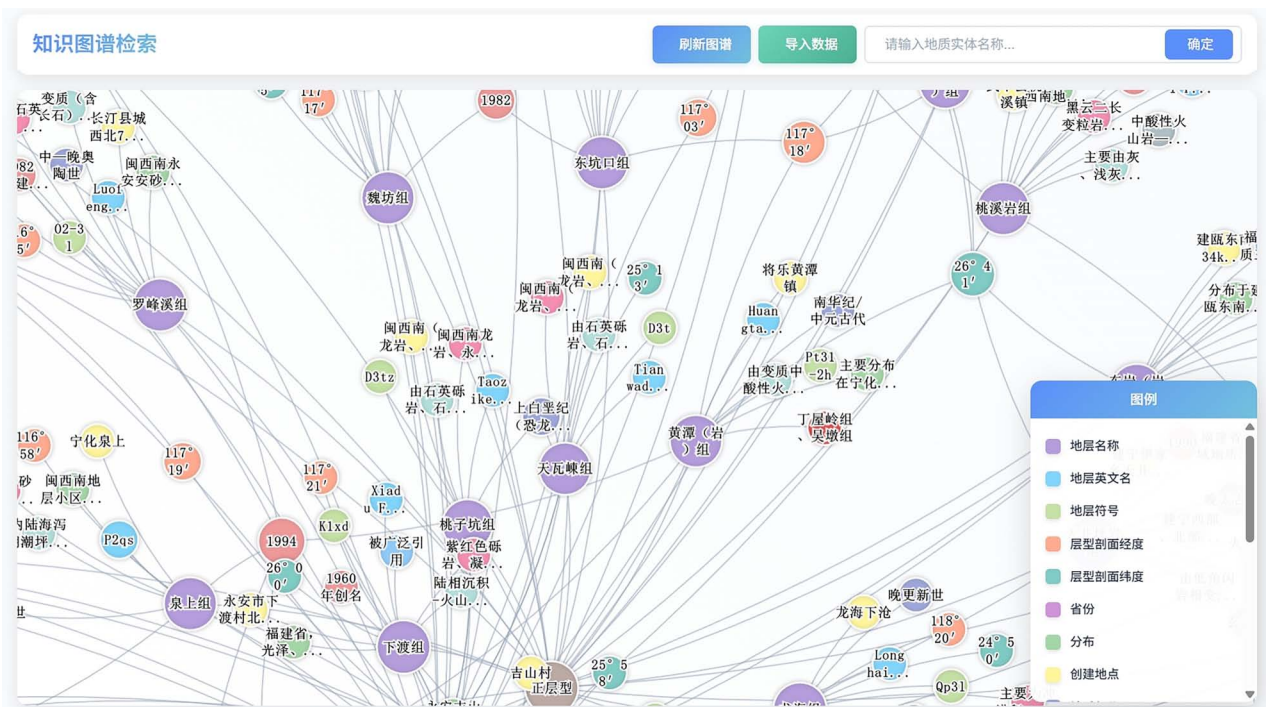


图 5 知识图谱可视化效果图  
Fig. 5 Knowledge graph visual rendering

directed Layout) 算法来自动组织图谱中节点的位置, 这种布局能够模拟物理世界中粒子间的引力和斥力, 使得相互关联紧密的节点自然聚集, 关系疏远的节点相互排斥, 从而形成一个结构清晰、易于观察

的网状图。具体效果如图 5 所示。

为了增强图谱的可读性, 系统设计了详细的图例。不同类型的地质实体 (如地层名称、地层符号、省份、分布等) 被赋予了不同的颜色, 用户可以通过

图例快速识别图中各个节点的类型。当用户将鼠标悬停在任意一个节点上时,系统会动态弹出一个信息框,清晰地展示该节点的实体名称、实体类型及其连接数(度),帮助用户快速获取节点的关键信息,提升了图谱的交互性和探索效率。

### 2.2.2 知识图谱检索

为满足用户针对特定实体进行深入分析的需求,本模块提供了高效的知识图谱检索功能。界面顶部设有检索框,用户可以输入感兴趣的地质实体名称。点击确定后,系统会以该实体为中心,高亮显示其自身以及所有直接与之相连的一度关系节点和连线,而将图谱中的其他无关节点进行隐藏或淡化处理。这种“聚焦”式的展示方式,能够帮助用户从复杂庞大的知识网络中迅速定位目标,并清晰地观察其直接关联的属性和关系,极大地提高了信息检索的效率和准确性。效果如图6所示。

### 2.3 多模态检索模块

为了弥补纯文本信息在直观性上的不足,并为地质研究提供更丰富的参考资料,系统设计了多模态检索模块。该模块旨在将结构化的地层文本描述与相关的地质图像资料进行关联,实现“文—图”互证的检索与浏览体验。具体效果如图7所示。

该模块的功能实现依赖于一个预先构建的多模态数据库。系统管理员可预先将整理好的地层资料存入数据库,每一条记录都包含地层的ID、名称、详

细的文本描述以及对应的图片URL。模块的主界面以列表形式清晰地展示了所有已录入的地层信息。用户可以通过浏览列表或利用页面顶部的检索框,输入地层名称等关键词进行快速筛选和定位。

对于列表中的任意地层条目,用户可以点击“查看详情”操作,进入该地层的多模态展示页面。该页面采用左右分栏的布局,左侧为该地层的详细文字描述,右侧则展示与之关联的地质剖面图等图片资料。这种并列呈现的方式,使用户能够方便地对照阅读,将抽象的文字描述与具象的视觉证据相结合,从而对地层特征形成更为全面和深刻的理解。该模块有效地整合了不同模态的地质数据,为用户提供了一个便捷的、集浏览与检索于一体的多模态资料库。

### 2.4 多特征关联分析模块

为支持更深层次的、探索性的文本分析,系统提供了多特征关联分析模块。该功能允许用户根据自定义的维度组合,对单篇地质文本中蕴含的复杂关系网络进行即时的可视化分析。与基于结构化数据库生成的知识图谱不同,此模块直接面向原始文本,动态地揭示实体间基于多种特征的内在联系。效果如图8所示。

用户首先在文本框中输入待分析的地质段落。随后,通过勾选一组预设的特征维度复选框(包括实体、关系、属性、时空、主题等),来定义本次分析

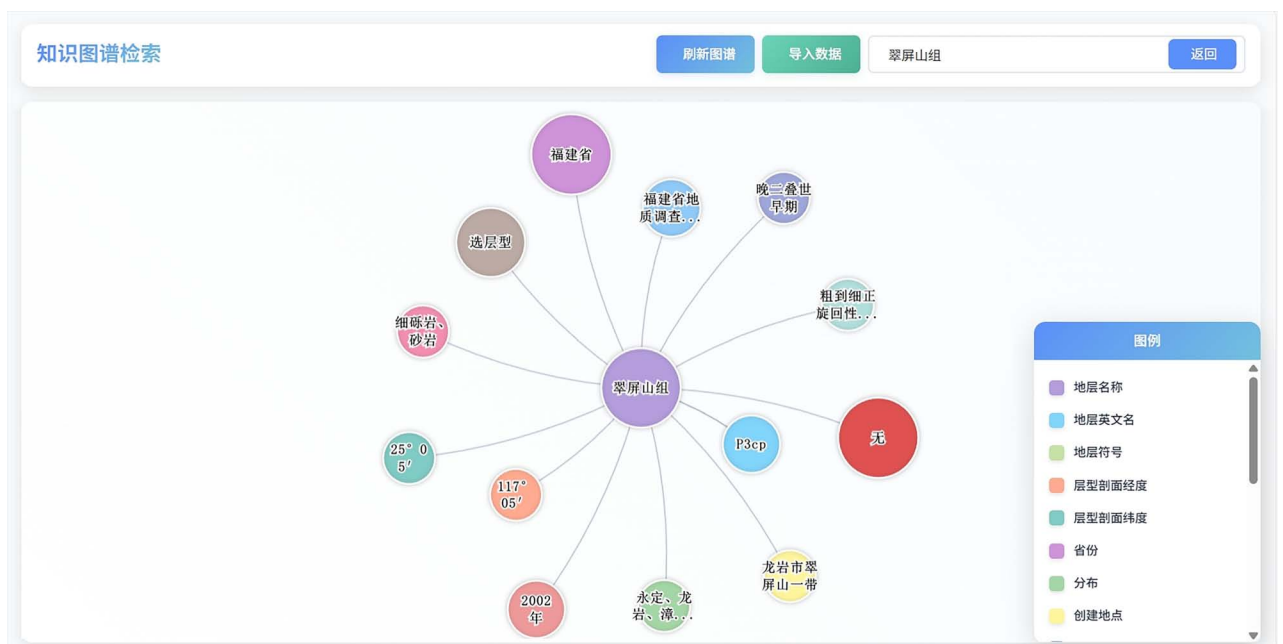


图6 知识图谱检索效果图

Fig. 6 Knowledge graph retrieval result visualization

### 天井坪岩组 (Pt1t)

文本

图片

天井坪岩组仅分布于闽西北建宁。为一套低角闪岩相变质岩系,由黑云(斜长)变粒岩、云母片岩组成,少量石英岩、斜长角闪岩,普遍含砂线石、铁铝榴石等特征变质矿物。混合岩化较强烈,脉体为花岗岩、长英质。原岩以砂泥质岩为主,夹基性、中性火山岩。天井坪岩组系福建省区域地质调查队(1990)所创“天井坪组”,命名地点位于建宁伊家乡天井坪村。付树超等(1991)介绍。原组一套韧性变形、混合岩化普遍,变质程度达低角闪岩相的变质岩系,原岩属火山复理石建造,时代归属为晚太古代。

20世纪70年代初,江西省区域地质调查大队在建宁地区调查时将相关的地质实体划分为震旦系下部与中部;福建省区域地质测量队(1977)划分为上、下震旦统;福建省地质矿产局(1985)亦划分为下震旦统丁屋岭组 and 上震旦统,同时认为建宁客坊的黑云钾长变粒岩原岩大致相当于石英角闪斑质火山岩。福建省闽西地质大队(1996)沿用天井坪组,在1:5万建宁县幅中部和北部划分出天井坪组。李兼海等(1997)沿用天井坪组,认为属现今福建省最老的地层,同时根据福建省区域地质调查队(1993)成果资料,将天井坪组分布限定在建宁县里心、客坊地区与浦城丘源一带,时代置晚太古代—古元古代。黄泉桢等(1998)、庄建民等(2000)均沿用天井坪组,时代置晚太古代,认为天井坪组分布局限,除分布于建宁西部里心、客坊地区外,尚分布于光泽司前积谷岭一带,而将浦城丘源一带归属为太古代(岩)组。福建省地质调查研究院(2004)将光泽司前积谷岭一带划属下降(岩)组,将浦城丘源一带归属为太古代(岩)组。江西省地质调查研究院(2007)根据“混合岩化”程度的强弱对建宁县里心、客坊等地区天井坪组进行解体,将花岗岩、长英质脉体含量大于60%的均划属晚奥陶世花岗岩类侵入体,将天井坪组时代置古元古代。

本志沿用庄建民等(2000)关于天井坪组定义,根据其变形改造特征等,采用构造岩石地层单位命名天井坪岩组。是指分布于建宁西部、北部里心、客坊、伊家地区的一套普遍含砂线石、铁铝榴石的黑云(斜长)变粒岩夹云母片岩、石英岩、少量斜长角闪岩,混合岩化较强烈,脉体为花岗岩、长英质。原岩以砂泥质岩为主,夹基性、中性火山岩。

天井坪岩组正层型剖面为建宁西南约18.5km伊家乡天井坪村剖面(116°44', 26°41') (图1-2), 分层描述如下: 古元古代天井坪岩组 (Pt1t) 总厚度 > 262.6m

9.深灰色黑云斜长角闪岩夹灰色薄层角闪斜长岩, 未见顶	> 6.9m
8.掩盖	32.8m
7.厚层黑云斜长变粒岩	50.2m
6.浅灰色厚层黑云斜长变粒岩与长石英英岩层, 顶部为深灰色厚层角闪斜长变粒岩	32.4m
5.灰黄色中层黑云变粒岩夹薄层绿帘角闪长石石英岩	36.6m
4.灰黄色中层黑云斜长黑云片岩	13.7m
3.灰黄色中层黑云砂线石云石英片岩夹薄层黑云斜长变粒岩	31.8m
2.薄层二云片岩夹灰绿色厚层黑云变粒岩, 局部见糜棱岩	36.5m
1.构造片岩、糜棱岩, 未见底	> 21.7m

建宁西部落马地—芝峰剖面, 是天井坪岩组岩性组合出露较全, 具有代表性的典型剖面(116°42', 26°52') (图1-3), 分层描述如下: 古元古代天井坪岩组 (Pt1t) 总厚度 > 855.3m

16.深灰色中厚层黑云斜长变粒岩夹二云片岩及透辉斜长石英岩, 未见顶	> 20.2m
15.黄棕色薄层白云片岩	27.0m
14.掩盖	205.1m
13.灰黄色薄层白云片岩	11.8m
12.灰黄色厚层斜长变粒岩(少量透辉角闪变粒岩)夹二云片岩及石英岩条带	7.2m

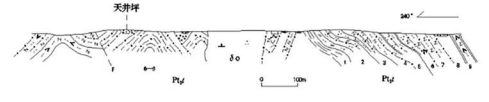


图1-2:建宁县天井坪组古元古代天井坪组实测剖面图

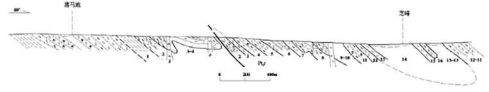


图1-3:建宁县落马地——芝峰古元古代天井坪组实测剖面图

图7 图文多模态详情展示界面

Fig. 7 Multimodal image-text detail display interface

导入文本

### 输入文字, 进行多特征关联

原岩属火山复理石建造, 时代归属为晚太古代。

20世纪70年代初, 江西省区域地质调查大队在建宁地区调查时将相关的地质实体划分为震旦系下部与中部;福建省区域地质测量队(1977)划分为上、下震旦统;福建省地质矿产局(1985)亦划分为下震旦统丁屋岭组和上震旦统,同时认为建宁客坊的黑云钾长变粒岩原岩大致相当于石英角闪斑质火山岩。福建省闽西地质大队(1996)沿用天井坪组,在1:5万建宁县幅中部和北部划分出天井坪组。李兼海等(1997)沿用天井坪组,认为属现今福建省最老的地层,同时根据福建省区域地质调查队(1993)成果资料,将天井坪组分布限定在建宁县里心、客坊地区与浦城丘源一带,时代置晚太古代—古元古代。黄泉桢等(1998)、庄建民等(2000)均沿用天井坪组,时代置晚太古代,认为天井坪组分布局限,除分布于建宁西部里心、客坊地区外,尚分布于光泽司前积谷岭一带,而将浦城丘源一带归属为太古代(岩)组。福建省地质调查研究院(2004)将光泽司前积谷岭一带划属下降(岩)组,将浦城丘源一带归属为太古代(岩)组。江西省地质调查研究院(2007)根据“混合岩化”程度的强弱对建宁县里心、客坊等地区天井坪组进行解体,将花岗岩、长英质脉体含量大于60%的均划属晚奥陶世花岗岩类侵入体,将天井坪组时代置古元古代。

确认

实体

关系

属性

时空

主题

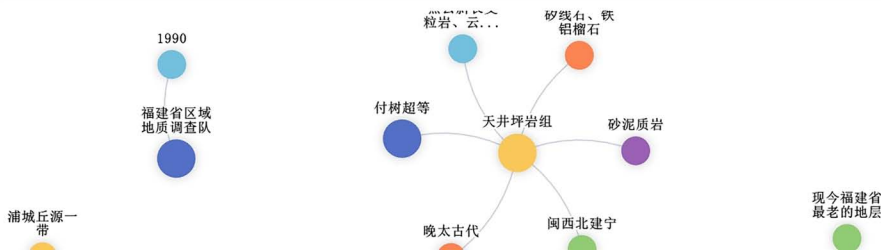


图8 多特征关联分析可视化图

Fig. 8 Multi-feature correlation analysis visualization

的关联视角。点击确认后,系统后端对文本进行快速的、多维度的信息抽取。例如,如果用户选择了“实体”和“属性”,系统会识别出文中的核心地质实体及其对应的属性信息。

分析完成后,结果会以一种“中心—辐射”状的图谱形式呈现出来。图谱的核心是文本中识别出的关键实体,而围绕这些核心实体辐射出的是根据用

户所选特征维度关联上的其他节点。例如,一个地层实体可能会连接到代表其岩性组成的“属性”节点,也可能连接到代表其分布区域的“时空”节点。这种高度定制化的、即时的可视化分析能力,为用户提供了一种强大的探索工具,能够灵活地从不同角度审视和挖掘文本内部的知识结构,发现潜在的、跨维度的信息组合与关联模式。

### 3 系统实现

本章将从代码层面,对系统的核心功能实现进行阐述,重点介绍前后端关键模块的逻辑与交互。

#### 3.1 后端实现

后端服务基于 Spring Boot 框架搭建,遵循经典的控制器、服务、数据访问三层架构模式,以实现业务逻辑的解耦和代码的可维护性。

项目核心控制器为 LithoController,它负责接收前端传递的 HTTP 请求。例如,信息抽取功能对应一个 POST 类型的 API 接口,即/api/litho。该接口通过 @RequestBody 注解接收前端发送的 JSON 对象,其中包含了待处理原始文本、用户选择的配置文件 ID 及调用的模型列表。

LithoService 是业务逻辑的核心。它首先根据 configId 从数据库中查询对应的抽取字段配置。接着,它会根据这些字段动态构建一个结构化的指令。这是与大语言模型交互的关键,一个设计良好的指令能够显著提升抽取结果的准确性。例如,指令会明确指示模型:“请从以下文本中,抽取出‘地层名称’、‘岩性特征’等信息,并以 JSON 格式返回。”

随后,服务通过 Spring 内置的 RestTemplate 工具,以 HTTP POST 请求的方式调用大语言模型服务 API。请求体中包含构建好的指令和原始文本。在收到模型的 JSON 响应后,后端服务会对其进行解析,将其转换为统一的数据结构对象,并最终返回给控制器。

数据持久化层采用 MyBatis-Plus,通过定义 Mapper 接口来操作数据库。例如,KnowledgeGraphMapper 负责将用户确认后的抽取结果或导入的 Excel 数据,分别存入实体表 entity 和关系表 relation 中,为知识图谱可视化提供数据源。

#### 3.2 前端实现

前端基于 Vue 3 和 Composition API 进行开发,利用 Vue Router 管理页面路由。核心页面组件包括 lithoextract.vue 和 KGragh.vue 等。

信息抽取功能主要由 lithoextract.vue 组件实现。该组件界面包含一个用于粘贴或导入文本的<textarea>输入框,一个包含配置文件下拉选择和模型多选框的配置面板,以及一个用于展示结果的<el-table>表格区域。当用户点击“开始抽取”按钮,会触发一个事件处理函数。此函数通过 Vue 的响应式系统,获取用户输入的文本和选择的配置。随后,它使用 axios 库将这些数据封装成一个 POST 请

求,发送至后端的/api/litho 接口。请求成功后,组件接收到后端返回的 JSON 格式的结构化数据,并将其赋值给 modelsData 的响应式变量。Vue 的数据绑定机制确保了当此变量更新时,页面上的表格会立即自动渲染出最新的抽取结果,无需手动操作 DOM。

知识图谱的可视化与检索功能则由 KGragh.vue 组件负责。此组件的核心是集成了强大的 ECharts 图表库,用于动态渲染和交互式操作图谱。在组件挂载完成后的 onMounted 生命周期钩子中,它会首先向后端请求初始的图谱数据,包括节点列表和关系列表。获取数据后,便初始化一个 ECharts 实例,并为其配置一个核心的 option 对象。其中,图表 series 类型被设为‘graph’,并采用‘force’力引导布局,使得节点能够根据它们间的关系自动分布,形成清晰的网络结构。节点和关系数据最终被赋给 series.data 和 series.links,并通过 setOption 方法完成图谱的绘制。该组件还实现了前端的实体检索功能。当用户在搜索框输入关键词,组件会根据关键词在本地已加载的全量数据中进行过滤,筛选出匹配的节点及其直接关联的节点和关系,然后使用这部分子数据重新渲染图谱,从而实现聚焦式的展示效果。

### 4 结论

本文围绕地质领域非结构化文本智能化抽取与可视化等问题,设计并实现了信息抽取与可视化系统。系统以 Spring Boot 与 Vue 为技术基础,将大语言模型引入地质信息抽取任务,并结合模板化配置,实现了非结构化文本到结构化数据的高效转化。在此基础上,系统进一步集成了知识图谱可视化、多模态检索和多特征关联分析等模块,使抽取结果不仅能以表格形式呈现,还能够通过图谱与图像等多维度方式进行直观展示与探索。整体上,该系统为地质文本的挖掘提供了一个可行的实现路径。

尽管取得了一定成效,本研究仍存在一些需要改进之处。首先,系统依赖通用大语言模型,在面对高度专业化的地质语料时仍可能出现理解偏差,未来可以结合抽取数据对模型进行领域微调,以提升精度与稳定性。其次,目前系统主要面向以地质志为主的单一数据源,后续可扩展至跨文档甚至跨区域多源数据,借助知识融合方法形成更大规模的时空知识网络。此外,在交互方式上,可以探索自然语言查询等更友好的交互手段,使用户能够通过问答

或语义检索的方式直接调用系统功能,从而进一步降低使用门槛。

## 参 考 文 献 / References

(The literature whose publishing year followed by a “&” is in Chinese with English abstract; The literature whose publishing year followed by a “#” is in Chinese without English abstract)

董家慧子, 谢忠, 邱芹军, 马凯, 田苗, 陶留锋. 2023. 融合容错机制的基于 Attention-Mask RCNN 地质表格信息抽取方法. 地质科学, 58(3): 1147~1163.

刘文聪, 张春菊, 汪陈, 张雪英, 朱月琴, 焦守涛, 鲁艳旭. 2021. 基于 BiLSTM-CRF 的中文地质时间信息抽取. 地球科学进展, 36(2): 211~220.

邱芹军, 段雨希, 田苗, 吴麒麟, 马凯, 陶留锋, 谢忠. 2024. 多模态数据的地质图关网络构建及知识服务. 地质论评, 70(4): 1469~1482.

邱芹军, 田苗, 马凯, 谢忠, 金相国, 段雨希, 陶留锋. 2023. 区域地质调查文本中文命名实体识别. 地质论评, 69(4): 1423~1433.

邱芹军, 田苗, 吴麒麟, 陈建国, 诸云强, 陈占龙, 谢忠. 2025. 基于多源异构数据的地质知识图谱构建与应用. 地学前缘, 1~17.

孙超, 黄浩. 2026. 基于 Gplates 和 Vue 的古地图在线可视化系统设计及实现. 地质论评, 72(1): 295~304.

王汉雨, 周永章, 许娅婷, 王维曦, 曹伟, 刘永强, 贺炬翔, 陆可飞. 2024. 基于微服务架构的城市土壤污染物物联网监测及可视化系统研发. 地学前缘, 31(4): 165~174.

王堃屹, 周永章. 2024. 粤西庞西垌地区非结构化地质信息机器可读表达与致矿异常区域智能预测. 地学前缘, 31(4): 47~57.

王远超, 赵元艺, 程贤达, 巩鑫. 2024. 上黑龙江盆地金属矿成矿系统. 地质论评, 70(S1): 63~64.

张泽琛, 张苏江, 刘晶晶, 肖淳, 高鹏鑫. 2026. 化探样品信息管理服务系统研究与实现. 地质论评, 72(1): 314~320.

周圣荃, 李以科, 王永志, 刘海明, 李楠, 柯昌辉, 李瑞萍, 赵永岗, 张丽. 2025. 生成式 AI 技术在地学研究中的应用现状及发展趋势. 地学前缘, 32(4): 303~316.

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, 杨威, 肖凡, 吉俊杰, 韩枫, 唐磊, 欧阳冲, 沈文杰. 2021. 软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.

Dagdelen J, Dunn A, Lee S, Walker N, Rosen A S, Ceder G, Persson K A, Jain A. 2024. Structured information extraction from scientific text with large language models. Nature Communications, 15: 1418.

Dong Jiahui, Xie Zhong, Qiu Qinjun, Ma Kai, Tian Miao, Tao Liufeng. 2023 #. Geological table information extraction method based on Attention-Mask RCNN with fault-tolerant mechanism. Chinese Journal of Geology, 58(3): 1147~1163.

Huang Zongcai, Peng Peng, Lu Feng, Zhang He. 2025. An LLM-based method for quality information extraction from web text for crowd-sensing spatiotemporal data. Transactions in GIS, 29(1): e13294.

Liu Wencong, Zhang Chunju, Wang Chen, Zhang Xueying, Zhu Yueqin, Jiao Shoutao, Lu Yanxu. 2021&. Geological time information extraction from Chinese text based on BiLSTM-CRF. Advances in Earth Science, 36(2): 211~220.

Qiu Qinjun, Duan Yuxi, Tian Miao, Wu Qirui, Ma Kai, Tao Liufeng,

Xie Zhong. 2024&. Geological map association network construction and knowledge service application for multimodal data. Geological Review, 70(4): 1469~1482.

Qiu Qinjun, Ma Kai, Lv Hairong, Tao Liufeng, Xie Zhong. 2023a. Construction and application of a knowledge graph for iron deposits using text mining analytics and a deep learning algorithm. Mathematical Geosciences, 55(3): 423~456.

Qiu Qinjun, Tian Miao, Ma Kai, Xie Zhong, Jin Xiangguo, Duan Yuxi, Tao Liufeng. 2023&. Chinese named entity recognition for regional geological survey text. Geological Review, 69(4): 1423~1433.

Qiu Qinjun, Tian Miao, Wu Qirui, Chen Jianguo, Zhu Yunqiang, Chen Zhanlong, Xie Zhong. 2025&. Construction and application of geological knowledge graph based on multi-source heterogeneous data. Earth Science Frontiers, 1~17.

Qiu Qinjun, Wang Bin, Ma Kai, Lü Hairong, Tao Liufeng, Xie Zhong. 2023b. A practical approach to constructing a geological knowledge graph: A case study of mineral exploration data. Journal of Earth Science, 34(5): 1374~1389.

Sun Chao, Huang Hao. 2026&. Design and implementation of an online paleogeographic map visualization system based on GPlates and Vue. Geological Review, 72(1): 295~304.

Wang Hanyu, Zhou Yongzhang, Xu Yating, Wang Weixi, Cao Wei, Liu Yongqiang, He Juxiang, Lu Kefei. 2024&. IoT monitoring and visualization of urban soil pollution based on mi-croservice architecture. Earth Science Frontiers, 31(4): 165~174.

Wang Kunyi, Zhou Yongzhang. 2024&. Machine-readable expression of unstructured geological information and intelligent prediction of mineralization associated anomaly areas in Pangxidong District, Guangdong, China. Earth Science Frontiers, 31(4): 47~57.

Wang Yuanchao, Zhao Yuanyi, Cheng Xianda, Gong Xin. 2024 #. Metallogenic system of metal deposits in the Upper Heilongjiang Basin. Geological Review, 70(S1): 63~64.

Zhang Jinbin, Zhu Jun, Guo Zhihao, Wu Jianlin, Guo Yukun, Lai Jianbo, Li Weilian. 2025. More intelligent knowledge graph: A large language model-driven method for knowledge representation in geospatial digital twins. International Journal of Applied Earth Observation and Geoinformation, 139: 104527.

Zhang Xueying, Huang Yi, Zhang Chunju, Ye Peng. 2022. Geoscience knowledge graph ( GeoKG ): Development, construction and challenges. Transactions in GIS, 26(6): 2480~2494.

Zhang Zechen, Zhang Sujiang, Liu Jingjing, Xiao Chun, Gao Pengxin. 2026&. Research and implementation of chemical exploration sample information management and service system. Geological Review, 72(1): 314~320.

Zhou Shengquan, Li Yike, Wang Yongzhi, Liu Haiming, Li Nan, Ke Changhui, Li Ruiping, Zhao Yonggang, Zhang Li. 2025&. Current status and development trends of generative AI technology in Earth science research. Earth Science Frontiers, 32(4): 303~316.

Zhou Yongzhang, Xiao Fan. 2024&. Overview: A glimpse of the latest advances in artificial intelligence and big data geoscience research. Earth Science Frontiers, 31(4): 1~6.

Zhou Yongzhang, Zhang Qianlong, Huang Yongjian, Yang Wei, Xiao Fan, Ji Junjie, Han Feng, Tang Lei, Ouyang Chong, Shen Wenjie. 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.

# Design and implementation of an intelligent extraction and visualization system for geological information via spring boot and vue

QIU Qinjun<sup>1)</sup>, LÜ Yunfeng<sup>1)</sup>, WU Qirui<sup>1)</sup>, TIAN Miao<sup>1)</sup>,  
SUN Jiangbing<sup>2, 3)</sup>, TAO Liufeng<sup>1)</sup>, ZHANG Yan<sup>3)</sup>

1) *School of Computer Science, China University of Geosciences, Wuhan, 430074;*

2) *School of Earth Sciences and Engineering, Hohai University, Nanjing, 211100;*

3) *Geological Data Archives of Jiangsu Province, Nanjing, 211100*

**Objectives:** Over the years, geological surveys and information infrastructure development have accumulated vast amounts of multi-source, heterogeneous geoscience data. However, traditional geological information processing relies on manual data entry and fragmented analysis, resulting in low data integration efficiency and delayed decision-making. To address the challenge of low utilization efficiency for unstructured geological texts, this study designed an intelligent information extraction and visualization system.

**Methods:** The system adopts a frontend-backend separation architecture. Multiple large language models are introduced as the core information extraction backbone. Domain knowledge is incorporated through user-defined templates to guide the extraction of structured information from multimodal geological data. Extracted geological entities and attributes are represented using representation learning techniques. Similarity measurement methods are applied to establish multi-feature associations among geological data, based on which a geological knowledge graph is constructed. Multidimensional interactive visualization is implemented using ECharts.

**Results:** This integrated platform achieves transformation from text to knowledge, automatically generating knowledge graphs to reveal entity relationships. It provides an effective solution for intelligent processing and deep utilization of geological texts. By integrating cutting-edge artificial intelligence and visualization technologies, it significantly enhances the usability and discovery efficiency of geological knowledge, laying a solid practical foundation for building an intelligent geological knowledge service platform.

**Conclusions:** This study focuses on intelligent extraction and visualization of unstructured geological texts and presents the design and implementation of an information extraction and visualization system for geological applications. Overall, the proposed system provides a feasible implementation pathway for the deep mining and utilization of geological textual data.

**Keywords:** geological knowledge graph; large language model; software tools; multimodal geological data

**Acknowledgements:** This work was supported by the National Natural Science Foundation of China (No. 42301492), the Open Fund of the Geological Data Intelligent Application Technology Innovation Center of Jiangsu Provincial Department of Natural Resources (No. GDIATIC-XM-202502), and the Geological Survey Project of the China Geological Survey (No. DD20240029)

**First author:** QIU Qinjun, male, born in 1988, Ph. D., associate professor, is mainly engaged in geological big data mining and geological knowledge graph research; Email: qiuqinjun@cug.edu.cn

**Corresponding author:** SUN Jiangbing, male, born in 1986, master's degree, senior engineer, is mainly engaged in comprehensive utilization of geological data, geological data mining and related application studies; Email: 230209020002@hhu.edu.cn

**Manuscript** received on: 2025-10-10; Accepted on: 2026-01-08; Published online on: 2026-02-15

**Doi:** 10.16509/j.georeview.2026.02.081

**Edited by:** LIU Zhiqiang

