



地震学报

ACTA SEISMOLOGICA SINICA



ISSN 0253-3782 CN 11-2021/P

大数据时代的地球科学知识图谱研究现状与展望

牛凤桂 张贝 陈石

Review and perspective of Earth Science Knowledge Graph in Big Data Era

Niu Fenggui, Zhang Bei, Chen Shi

引用本文:

牛凤桂, 张贝, 陈石. 2024. 大数据时代的地球科学知识图谱研究现状与展望. *地震学报*[J], 46(3): 353–376.

DOI: 10.11939/jass.20230157

Niu F G, Zhang B, Chen S. 2024. Review and perspective of Earth Science Knowledge Graph in Big Data Era. *Acta Seismologica Sinica*[J], 46(3): 353–376. DOI: [10.11939/jass.20230157](https://doi.org/10.11939/jass.20230157)

在线阅读 View online: <https://doi.org/10.11939/jass.20230157>

您可能感兴趣的其他文章

Articles you may be interested in

面向地震烈度仪的现地地震动预测模型的构建与验证

Construction and verification of onsite ground motion prediction models for seismic intensity instrument
地震学报. 2021, 43(5): 643–655 <https://doi.org/10.11939/jass.20210075>

郟庐断裂带江苏段新生界三维地质构造建模

Three-dimensional geological modeling of Cenozoic erathem in Jiangsu segment of the Tanlu fault zone
地震学报. 2020, 42(2): 216–230 <https://doi.org/10.11939/jass.20190131>

巴彦浩特地震台石英水平摆倾斜仪观测资料分析

Analyses on observation data of quartz horizontal pendulum tiltmeter at the seismic station Bayanhot, Inner Mongolia
地震学报. 2020, 42(1): 66–78 <https://doi.org/10.11939/jass.20180137>

龙门山断裂带地壳结构的三维建模

Three-dimensional modeling of the crust structure of Longmenshan fault zone
地震学报. 2019, 41(6): 743–756 <https://doi.org/10.11939/jass.20190087>

用最小二乘配置法构建局部重力场模型

Regional gravity field model constructed by the least squares collocation
地震学报. 2020, 42(1): 53–65 <https://doi.org/10.11939/jass.20190084>

延庆台地电阻率铁质干扰的有限元模拟

The finite element modelling of iron interference on earth resistivity of Yanqing station
地震学报. 2017, 39(4): 520–530 <https://doi.org/10.11939/jass.2017.04.007>

牛凤桂, 张贝, 陈石. 2024. 大数据时代的地球科学知识图谱研究现状与展望. 地震学报, 46(3): 353-376. doi: 10.11939/jass.20230157.

Niu F G, Zhang B, Chen S. 2024. Review and perspective of Earth Science Knowledge Graph in Big Data Era. *Acta Seismologica Sinica*, 46(3): 353-376. doi: 10.11939/jass.20230157.



大数据时代的地球科学知识图谱 研究现状与展望*

牛凤桂^{1,2)} 张贝^{1,2)} 陈石^{1,2),†}

1) 中国北京 100081 中国地震局地球物理研究所

2) 中国北京 100095 北京白家疃地球科学国家野外科学观测研究站

摘要 地球科学是一门数据密集型学科, 而知识图谱则被认为是充分利用和发挥地球科学领域庞大数据的一种有效方式. 相对于通用大模型技术的发展, 知识图谱能提供更为准确的知识, 增强生成式大模型结果的智能性和可靠性. 本文首先阐释了知识图谱的相关概念和构建方法; 然后, 对在科学领域广泛应用的顶层本体基本形式化本体模型进行了详细介绍, 简单总结了地球科学领域已经建成的知识图谱, 并着重论述了地球科学领域的地质学核心本体和地质科学本体, 以及两者的异同; 最后, 介绍了与地球科学知识图谱构建相关的深时数字地球计划等国际地学数据科学计划, 并对未来地学知识图谱发展和地震科学知识图谱构建发展所面临的挑战和应用前景进行了展望.

关键词 知识图谱 本体构建 地球科学 地质本体

doi: 10.11939/jass.20230157 中图分类号: P315.09 文献标识码: A

Review and perspective of Earth Science Knowledge Graph in Big Data Era

Niu Fenggui^{1,2)} Zhang Bei^{1,2)} Chen Shi^{1,2),†}

1) *Institute of Geophysics, China Earthquake Administration, Beijing 100081, China*

2) *Beijing Baijiatuan Earth Sciences National Observation and Research Station, Beijing 100095, China*

Abstract: Earth Science is a discipline that heavily relies on data, yet it is not fully harnessing the advantages of Earth data with existing technological means though covers many subject areas, Knowledge Graphs (KGs) is widely recognized as an effective approach to fully harness

* 基金项目 国家自然科学基金地震联合基金(U1939205)和中国地震局地球物理研究所基本科研业务费专项(DQJB22X11)联合资助.

收稿日期 2023-12-06 收到初稿, 2024-03-16 决定采用修改稿.

作者简介 牛凤桂, 在读硕士研究生, 主要从事地震科学数据分析和应用研究, e-mail: niufenggui@cea-igp.ac.cn; † 通信作者: 陈石, 博士, 研究员, 主要从事重力学与地球动力学相关研究, e-mail: chenshi@cea-igp.ac.cn



and utilize the extensive data in this field. Earth Science Knowledge Graphs can integrate geoscience knowledge, enhance research efficiency, and facilitate interdisciplinary collaboration. By analyzing network connections and semantic relationships, they uncover knowledge associations and patterns, and aid researchers in identifying new domains and posing novel research questions. Unlike conventional advancements in large-scale modeling technologies, Knowledge Graph offers precise knowledge that enhances both the intelligence and dependability of generated outcomes from such models.

Firstly, this study provides a detailed exposition of Knowledge Graph concepts and construction methods. Knowledge Graphs, as a form of data graph, are designed to collect and convey knowledge from the real world. Their universal expression is in the form of triples, consisting of head entities, tail entities, and the relationships between them. Knowledge Graphs have emerged as a significant approach for organizing structured knowledge and integrating information from multiple data sources in the organizational world. Their architectural framework primarily encompasses four components: source data acquisition, knowledge fusion, knowledge computation, and knowledge application. Source data acquisition stands as the primary step in building Knowledge Graphs, focusing on extracting useful information from various types of data. Knowledge fusion is pivotal in addressing the heterogeneity of different Knowledge Graphs, with the aim of enhancing their quality through integration. Knowledge computation represents the primary output capability of Knowledge Graphs, currently applied in fields such as semantic search, question answering, and visualization analysis. Knowledge Graph construction technology enables the extraction of information from structured, unstructured, and semi-structured data sources, organizing this information into knowledge and presenting it in graphical form. Presently, the construction of Knowledge Graphs in the field of Earth Sciences primarily employs two methods: Top-down and bottom-up approaches, with the overarching principle being the synthesis of both methods while allowing flexibility in their specific sequencing.

Secondly, this study offers a comprehensive introduction to the widely applied top-level ontology, the Basic Formal Ontology (BFO) model, in the scientific domain. The paper briefly summarizes existing Knowledge Graph in the geoscience field, emphasizing the GeoCore Ontology and Geoscience ontology (GSO) in the Earth Science domain, highlighting their similarities and differences. BFO, comprising 38 classes, is designed to facilitate information integration, retrieval, and analysis in scientific research. Presently, BFO has been successfully employed in over 350 ontology projects worldwide. The GeoCore Ontology, built upon BFO, serves as a specialized framework to describe the core concepts within the domain of Earth Science, rigorously defining a set of universal geological concepts during its development. Conversely, GSO provides a systematic framework for representing crucial geological science knowledge, encompassing three hierarchical layers: foundational, geological, and detailed modules. GeoCore can be viewed as an intermediary layer within GSO, which can be further expanded, while detailed modules have already been constructed within GSO. Additionally, researchers worldwide employ various methods such as literature mining, domain expert interviews, and data mining techniques to extract Earth Science knowledge from relevant literature, databases, and open data, subsequently to construct Knowledge Graphs. These Knowledge Graphs are found in applications across various domains including geological exploration, natural disaster prediction, and environmental conservation, and are utilized in practical projects

such as oil and gas exploration, water resource management, and climate change research. In summary, the application scope of Earth Science Knowledge Graphs is extensive, providing a crucial foundation of data and knowledge for scientific research, decision support, and sustainable development.

Finally, the study introduces international Earth Science data science initiatives such as the Deep-time Digital Earth (DDE) project related to constructing Earth Science Knowledge Graph, and the challenges and application prospects for the future development of Earth Science Knowledge Graph, with a focus on seismic science. The DDE aims to connect and coordinate global deep-earth data, promoting the sharing of geoscientific knowledge worldwide and facilitating research on Earth's evolution in a data-driven manner. Apart from the DDE, numerous domestic and international organizations and initiatives are driving the development of Knowledge Graph in Earth Science, such as OneGeology, EarthCube, and LinkedGeoData projects. Despite facing various challenges, Knowledge Graph is gradually overcoming these hurdles with advancements in technology and tools. These challenges are not exclusive to the field of Earth Science but are prevalent across all Knowledge Graph construction endeavors. However, due to the complexity and diversity of Earth Science, Knowledge Graph construction in this field encounters unique difficulties. Nevertheless, there is ample room for the creation and application of Knowledge Graph in Earth Science, with the introduction of Large Language Models (LLMs) bringing forth new opportunities. Earthquake Science, as a crucial branch of Earth Science, encompasses intersections of multiple primary disciplines such as geology, geophysics, and engineering seismology. However, the application of Knowledge Graphs in the field of Earthquake Science still faces significant gaps and urgently requires further research building upon existing models. In conclusion, the future development of Earth Science Knowledge Graphs will be an ongoing process of evolution and refinement, bringing more opportunities and benefits for fields such as Earth Science research, decision-making, and public education through sustained technological innovation and interdisciplinary collaboration.

Key words: Knowledge Graph; ontology construction; Earth science; geological ontology

引言

OpenAI 发布了 Chat Generative Pre-trained Transformer (简称为 ChatGPT), 彻底改变了人工智能与人机交互的方式 (Kocoń *et al.*, 2023), 推动了人工智能由深度学习向大模型时代的转变. 但是如果将通用语言大模型应用到专业领域, 对于事实准确性、可信度、可追溯等方面问题还有待突破. 而基于专业领域知识构建的知识图谱就是有效解决方式之一, 知识图谱 (Chen *et al.*, 2023) 可为人工智能大模型提供更为准确的知识来源, 用于提高生成式人工智能其结果的准确性, 以胜任专业的关键任务场景, ChatGPT 也可以帮助我们开发和管理知识图谱 (Meyer *et al.*, 2023).

地球科学是一门研究地球的科学, 涵盖了地理学、地质学、地球化学、地球物理学、大气科学和海洋科学等多个学科领域 (Zhang *et al.*, 2022), 每个学科都有庞大的知识体系. 由于地球各个方面的相互作用, 地球科学的知识与其它学科的知识有着密切的联系, 需要运用不同学科的理论和方法. 此外, 地球科学的研究离不开多种数据来源的支撑, 随着观测技术的不断发展, 地球科学数据的获取渠道日益丰富. 面对这些来自不同来源的数据, 研究者需要

进行数据处理与整合,考虑数据的准确性和可信度,并针对不同的空间和时间尺度的现象与过程进行研究和分析.由于地球是一个复杂的系统,其内部过程、相互作用和反馈机制多样,且研究对象常常伴随不确定性和变动性,使得地球科学知识的组织、检索和应用较为困难.

由于现有的技术手段不能充分利用和发挥地球数据的优势,在数据挖掘、数据集成、知识发现等各个领域存在巨大挑战(Zhou *et al.*, 2021),若想要系统地了解和获取到全面的研究进展和前沿知识,则需要一种新的、更有效的地球科学知识组织和检索方法来解决这些问题.近年来,机器学习和深度学习在地质学领域的应用显著增长,人工智能在地质学领域受到越来越多的关注(Ma, 2022).知识图谱作为一种新兴的知识表示和组织方式,具有知识关联、语义化、可扩展性、丰富的推理和灵活的应用等特点,可以弥补现有方法的不足.

在传统的图谱模式基础上,地球科学知识图谱扩展了地球科学独特的时空特征,并结合地图、文本、数字等地质学知识元素,建立了全领域地球科学知识表现模型,地球科学知识图谱不仅深化了地质学大数据分析,还促进了地质时间尺度构建、智能地图编制、地球科学知识演化与推理分析等领域的发展(Zhou *et al.*, 2021),具有广泛的研究价值和应用潜力.此外,地球科学知识图谱可以整合各种地球科学知识,提高研究效率,促进不同学科之间的交叉研究,也能够通过分析网络连接和语义关系,发现潜在的知识关联和规律,帮助研究者发现新的领域,提出新的研究问题.如果将专业领域知识图谱接入类似 ChatGPT 的通用大模型,则可以更好地识别对话中的实体及其之间的关系,减少歧义的产生,使得通用人工智能更容易理解专业领域问题从而生成更合适的答案.

本文总结了知识图谱在地球科学领域的构建与应用进展,旨在回顾目前已有的研究成果,并展望未来发展方向.首先介绍了知识图谱的概念及构建方法,然后重点介绍了各子学科领域已建立的知识图谱,尤其是地质学领域的应用.接着讨论了相关项目和组织,重点探讨了深时数字地球计划的工作.最后简要探讨了知识图谱面临的挑战以及对未来的展望.

1 知识图谱

1.1 知识图谱的定义与概念

知识图谱的概念由 Google 公司在 2012 年首次提出,包含超过 5 亿个实体和 7 亿个事实(Wang *et al.*, 2019),所涉及的基本要素可以追溯到古代历史,其核心思想是以图解的形式表示知识(Gutiérrez, Sequeda, 2021).尽管概念比较新,但是知识图谱并非一个全新的研究领域,在 2006 年 Berners-Lee 提出数据链接(linked data)的思想之后,语义网络研究掀起了一股热潮,而知识图谱技术建立在相关研究成果的基础之上,是对现有语义网络技术的进一步改进和提升(刘峤等, 2016).现如今,知识图谱的定义仍然是存在争议的,出现了很多不同的定义,甚至有些定义彼此之间相互冲突.此处,我们将知识图谱描述为一个数据图,其目的是积累和传递真实世界的知识.图中的节点表示感兴趣的实体,而边表示这些实体之间的潜在关系(Hogan *et al.*, 2021).知识图谱通过将实体和关系的信息以结构化的方式进行表示,能够帮助机器理解和推理关于这些实体和关系的语义信息.

数字字典仅提供了各种知识点的术语解释,但不包含它们之间的语义关系,因此无法进行语义翻译和推理,也无法被机器所理解.知识体系覆盖了特定领域内的知识点和它们之间的关系,可进行语义翻译,但描述形式通常为自然语言,难以被机器所理解.知识图谱以知识体系为基础,使用本体建模语言编码,形成机器可理解的模式,支持机器进行语义翻译和

数据融合(齐浩等, 2020).

用于存储的知识表示模型有多种, 包括资源描述框架(resource description framework, 缩写为 RDF)、资源描述框架模式(resource description framework schema, 缩写为 RDFS)、JavaScript 对象表示法(JavaScript object notation, 缩写为 JSON)等(McBride, 2004; Bassett, 2015; Tiwari *et al*, 2021)(表 1). 知识图谱的一种通用表达形式是三元组形式, 即头实体、尾实体和两个实体之间的关系(张吉祥等, 2022). 图 1 显示了通用知识图谱中实体和关系的一个示例, 例如事实“蒂姆·伯纳斯·李发明了万维网”包含两个实体, 即“蒂姆·伯纳斯·李”和“万维网”, 关系为“发明”, 形成了三元组“蒂姆·伯纳斯·李”、“发明”、“万维网”.

表 1 术语信息列表

Table 1 List of terminology information

名称	英文全称及缩写	意义
知识图谱	Knowledge Graph, 缩写为KG	数据图, 目的是积累和传递真实世界的知识, 在图中, 节点表示感兴趣的实体, 而边表示这些实体之间的潜在不同关系.
资源描述框架	Resource Description Framework, 缩写为RDF	最常用的模型, 以一组概念、关系和事实来组织知识, 这些概念、关系和事实由两种类型{实体、关系、实体}和{实体、属性、属性值}相关联.
资源描述框架模式	Resource Description Framework Schema, 缩写为RDFS	RDF词汇描述语言, 描述了资源的属性、类别以及它们之间的关系.
JavaScript对象表示法	JavaScript Object Notation, 缩写为JSON	一种数据交换格式, 与编程语言无关, 基于JavaScript的对象文字注释, 以一种对通用编程概念友好的方式表示数据.
基本形式化本体	Basic Formal Ontology, 缩写为BFO	由38个类组成的顶层本体, 旨在支持科学研究中各个领域的信息集成、检索和分析.
语言和认知工程的描述性本体论	Descriptive Ontology for Linguistic and Cognitive Engineering, 缩写为DOLCE	基于形式本体论的描述性本体, 它以一种准确和可靠的方式为语言和认知工程提供共享的语义基础, 通过定义和描述事物之间的关系和属性, 使得计算机能够更好地理解和处理自然语言和认知概念.
统一基础本体	Unified Foundational Ontology, 缩写为UFO	近十年来在哲学、认知科学、语言学 and 哲学逻辑等领域基于形式本体论整合发展起来的, 包含了多组关于基本概念建模的微观理论, 包括实体类型和关系类型.
深时数字地球计划	Deep-time Digital Earth program, 缩写为DDE	链接和协调全球深层地球数据, 共享全球地学知识, 旨在通过深层时间刺激数据驱动的发展来研究地球的演化.

知识图谱已经成为一种重要方法, 它能组织和整合结构化知识, 以及从多个数据源中提取的信息(Chaudhri *et al*, 2022). Chen 等(2021)基于 1991—2020 年发表的 386 篇科研论文的研究发现, 对知识图谱的研究兴趣在此期间明显上升, 并且不断扩大.

1.2 知识图谱的体系架构

知识图谱的体系架构主要分为三个部分(图 2): 源数据的获取、知识融合、知识计算和知识应用(张吉祥等, 2022).

源数据获取是构建知识图谱的第一步, 其关键在于如何从各种类型的数据中提取有用信息. 设计的技术包括命名实体识别、关系抽取和实体链接. 命名实体识别, 又称为实体抽取, 指的是自动从文本数据集中识别出命名实体. 经过命名实体识别后, 得到一系列离散的命名实体; 关系抽取需获取语义信息, 从相关语料中提取实体之间的关联关系, 并通过这些关系将实体连接起来, 以形成网状的知识结构; 实体链接则是将从文本中抽取得到的实体对象, 与知识库中对应的正确实体对象进行关联的操作.

通过自然语言处理技术, 已经成功地从非结构化、半结构化和结构化数据中提取出实体、关系和属性, 但这些信息仍需进一步清理和整合. 而知识融合是解决不同知识图谱异构

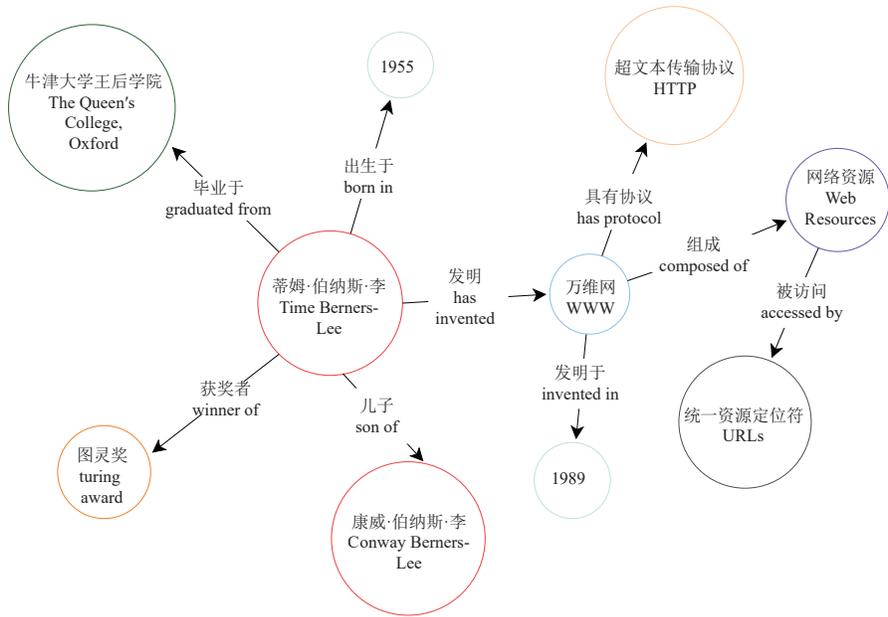


图 1 知识图谱中实体和关系的示例(修改自 Abu-Salih, 2021)

圈内为实体, 箭头为关系

Fig. 1 Examples of entities and relationships in Knowledge Graph (modified from Abu-Salih, 2021)

The in-circle contents are entities and the arrows are relationships

问题的关键, 通过知识融合, 我们能够提高知识图谱的质量。

知识计算是知识图谱的主要输出能力, 其中知识推理是其中最为重要的能力之一。通过知识推理, 能够从现有的知识中发现新的知识。知识图谱提供了一种更好地组织、管理和利用海量信息的方式, 目前, 知识图谱技术主要应用于语义搜索、问答和可视化分析等领域。

1.3 知识图谱的构建

KG 构建技术可以从结构化、非结构化甚至半结构化的数据源中提取信息, 并最终将这些信息组织成知识, 以图的形式表示出来 (Tiwari et al, 2021)。Abu-Salih (2021) 根据知识抽取的层次、知识库的类型以及集成的构建方法对 KG 的构建进行了分类 (图 3)。

地球科学知识图谱的构建目前主要包含两种方法: 自顶向下和自底向上。

自顶向下的方法是指通过手动输入地球科学专家的知识来构建知识图谱, 这种方法多用于专业领域的知识图谱。首先需要手动设计一个数据模式框架, 从顶层开始填充概念知识, 并且逐渐细化, 形成结构化和分层次的骨架, 然后按照顺序进行知识填充。这是一个将本领域专家大脑中的知识转化为机器可读并加以表示的过程。现有的很多地球科学知识图谱都是通过这种方法构建的, 比如 SWEET 本体 (Raskin, Pan, 2005)、GeoCore 本体 (Garcia et al, 2020) 等。这种方法的模式比较固定, 但是具有高精度的数据集 (Hogan et al, 2021; Zhang et al, 2022)。

自底向上的方法是指通过人工智能技术从地球大数据中构建知识图谱, 例如从地球科学文献中提取信息, 并对其重新分类和整理, 以构建相应的知识图谱。这种方法模型多变, 能够处理大量数据集并快速构建大型知识图谱, 但是自动提取知识的可靠性不高, 需要

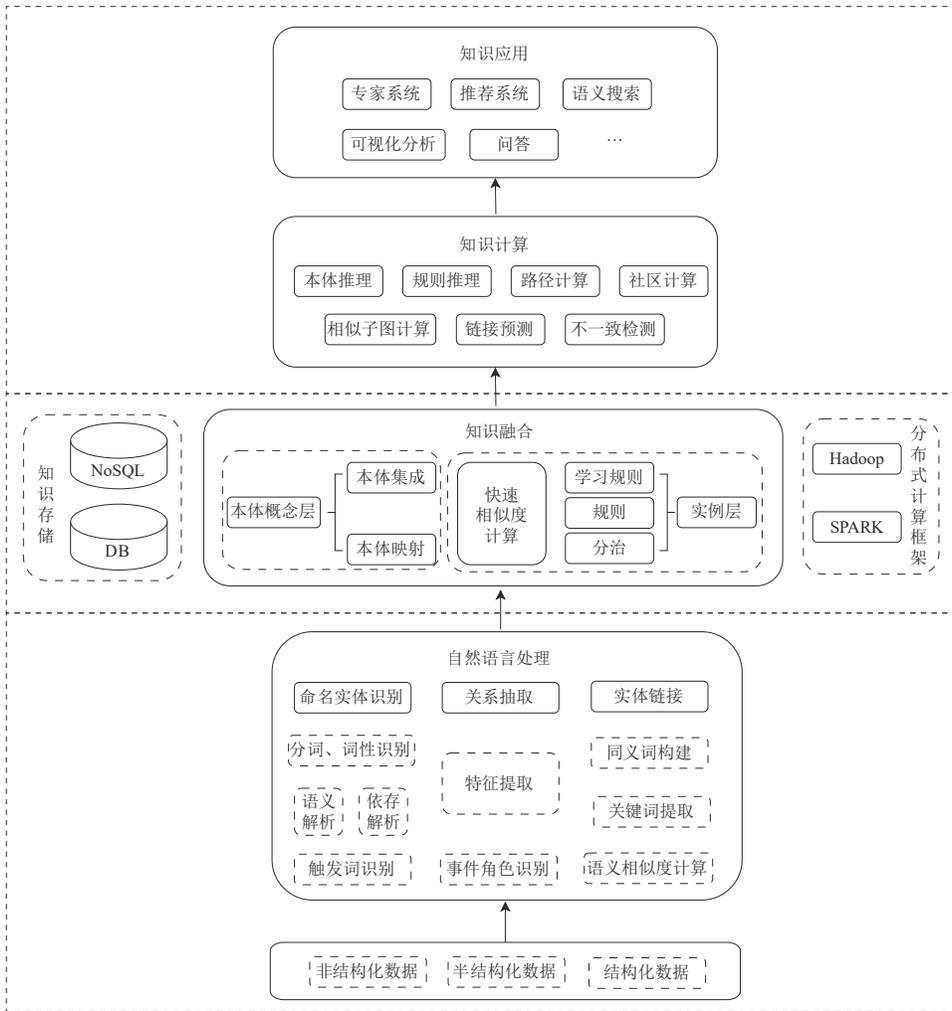


图 2 知识图谱的体系架构(修改自张吉祥等, 2022)

Fig. 2 The architectural framework of knowledge graph (modified from Zhang et al, 2022)

专家的介入来验证构建的知识图谱是否准确(Hogan et al, 2021; Zhang et al, 2022).

然而, 单纯只使用上述两种方法中的其中一个不足以构建完整的地球科学知识体系. 一方面, 地球科学中有很多专业知识具有高度不确定性或歧义, 很难建立统一的标准; 另一方面, 虽然地球大数据蕴含着丰富的地球科学知识, 但自动构建的知识体系并不完整, 例如在生成的知识图谱中, 实体和关系的精确逻辑表示和断言方面, 专家经验至关重要. 因此, 将地球科学领域专家的经验知识和计算机系统相结合的方法是相对较好的, 既可以充分发挥专家的优势, 也可以利用不断优化的人工智能方法来促进地球科学知识图谱的未来发展(Hogan et al, 2021; 张吉祥等, 2022). 地球科学领域专家提供的标注数据越多, 知识图谱的性能就越好(Zhang et al, 2022).

Ma (2022)总结了知识图谱建设的工作流程和建议, 并提出了在地球科学中构建和应用知识图谱的建议工作流程, 主题思想是自顶向下和自底向上两种方法相结合, 但是每一种方法的先后顺序可以改变(图 4).

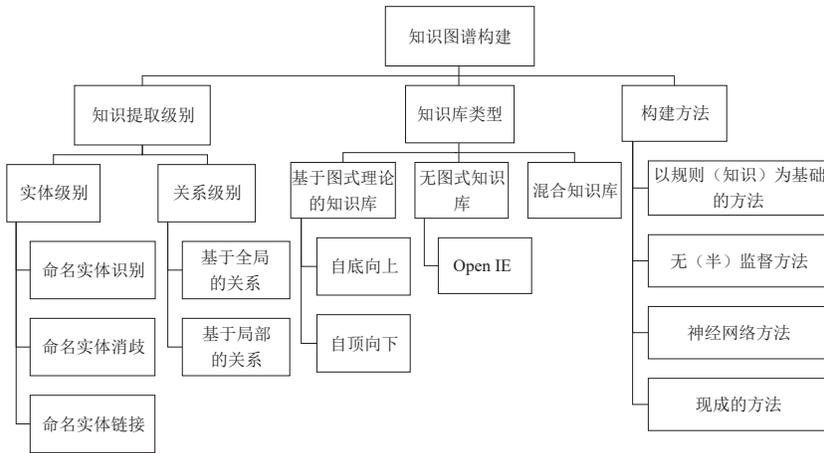


图 3 知识图谱构建方法的分类 (修改自 Abu-Salih, 2021)

Fig. 3 Classification of Knowledge Graph construction methods (modified from Abu-Salih, 2021)

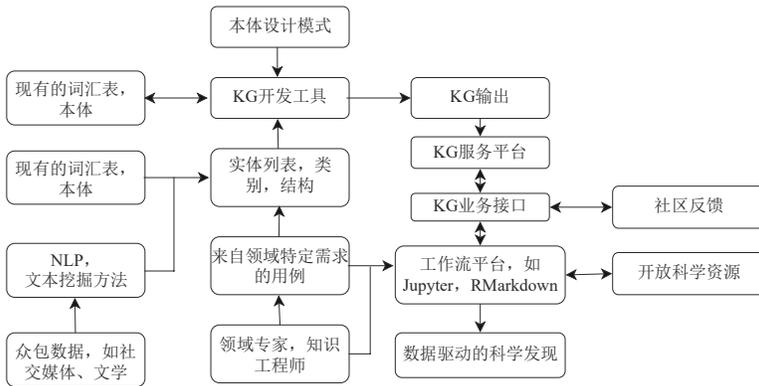


图 4 地球科学构建和实现知识图谱的工作流(修改自 Ma, 2022)

Fig. 4 Earth science workflow for constructing and implementing Knowledge Graphs (modified from Ma, 2022)

第一种形式是先使用自顶向下的方法，确定一个领域的主题，根据数据集确定适合的算法，并且寻求相关研究方向的研究人员的帮助。领域专家与知识工程师一起来分析每一个实例，获得实体、属性和关系之间的初步结构框架。在建立知识图谱过程中，可以使用自底向上的方法来扩充列表，并且选取一部分实例来测试知识图谱。在此过程中，可以重复使用和改编已经存在的本体，比如 SWEET 本体就是基于以上方法建立的。

第二种形式是先采用自底向上的方法从大量积累的数据中获取知识，通过大数据分析可能会获得人类现有的专业知识之外的一些规则，帮助我们发现新的模型和方法。接下来再采用自顶向下的方法，将相关研究方向的研究人员聚集到一起，对所得到的结果进行验证，中间过程可以利用现有的社区标准和本体模式，减少知识图谱中的不一致和重复性工作，达到提高知识图谱的质量和精度的目的。

在知识图谱的构建过程中，为了准确描述其中的概念、关系和属性，需要一个统一的、精确的语义框架。基于这样的需求，基本形式化本体作为一种被广泛接受的本体模型被引入和应用。通过使用基本形式化本体，我们能够建立起一个明确的概念体系，进而更好地理解解

和组织知识. 这种方法使知识图谱的构建和应用更高效.

1.4 基本形式化本体 (BFO)

本体是对共享概念的正式、明确的规范(Studer *et al*, 1998). 根据本体的概括性程度可以分为顶层本体(上层本体、基础本体)、领域本体和核心本体(Guarino, 1998). 顶层本体是描述空间、时间等共同的一般概念, 领域本体是专门描述某一学科领域, 核心本体处于顶层本体和领域本体之间(Oberle, 2006), 可以作为开发领域本体的基准(Scherp *et al*, 2011). 目前, 除了通用且开放的知识图谱, 如 Google KG (Abu-Salih, 2021), 知识图谱大多是与领域相关的, 并且在设计上通常包含特定领域的底层本体.

基本形式本体(basic formal ontology, 缩写为 BFO)是一个由 38 个类组成的顶层本体, 旨在支持科学研究中各个领域的信息集成、检索和分析. 目前, BFO 已经成功应用于全球 350 多个本体项目. BFO 是一个真正的顶层本体, 不包含特定于物理、医学、心理学等物质领域的术语(Otte *et al*, 2022). BFO 可以帮助领域专家更好地组织和表示领域知识, 促进不同计算机系统间的互操作性和推理能力, 并且已经在医学、生物学、工程等领域得到了广泛的应用.

中国的本体学者成立了本体中国(OntoChina)社区, 它是一个开放交流平台, 旨在促进本体技术的发展与应用, 其中就包括对 BFO 的介绍. 朱彦等(2021)系统地介绍了 BFO 的来源、特点, 将 BFO 翻译成中文, 并且简要梳理了层次结构, 如图 5 所示.

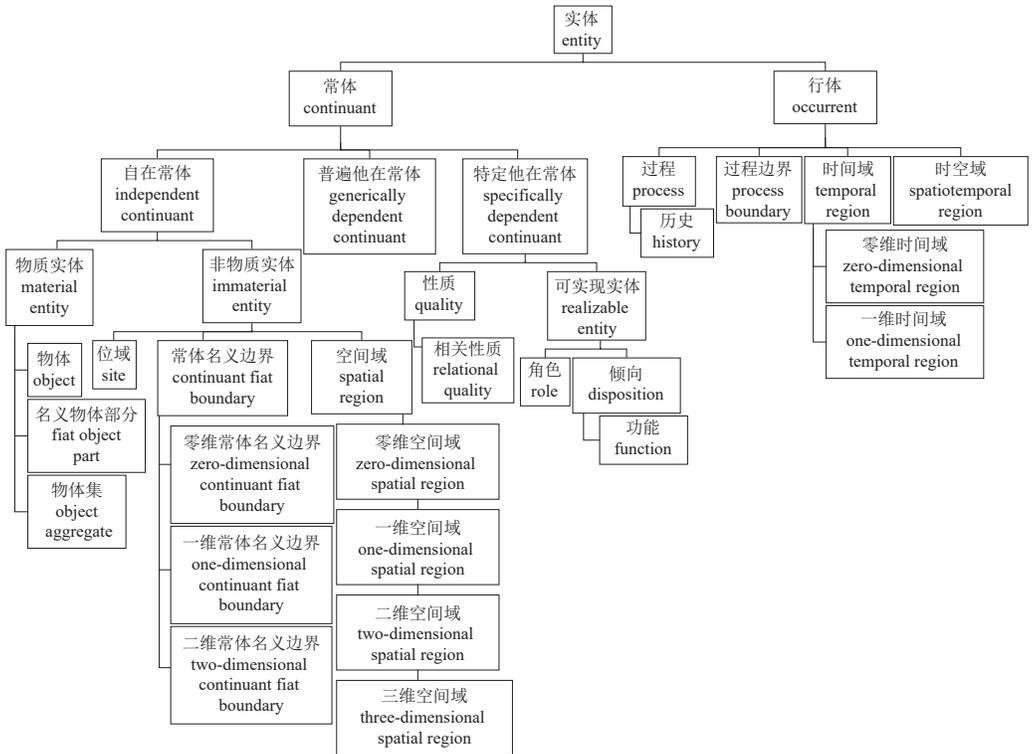


图 5 BFO 类层次结构(修改自朱彦等, 2021)

Fig. 5 BFO class hierarchy (modified from Zhu *et al*, 2021)

2 地球科学领域的本体模型研究

BabelNet, YAGO, Cyc, NELL, CaLiGraph 和 DBpedia 知识库是一些稳定演化的开放世界知识图谱。尽管通用的和开放世界的知识图谱广泛应用于处理不同领域不相关任务,但是构建领域特定的知识图谱以解决领域特定问题至关重要。领域知识图谱明确地概括了以与语义相关的实体和关系为代表的高层次主题领域及其特定子领域的概念(Abu-Salih, 2021)。

国内外研究者通过文献挖掘、领域专家采访、数据挖掘技术等手段,从地球科学相关文献、数据库和开放数据中提取地球科学知识,并构建知识图谱,应用于地质勘探、自然灾害预测、环境保护等领域(毛文山等, 2019; Fan *et al.*, 2019; 杜志强等, 2020; 邱芹军等, 2023; Liu *et al.*, 2023), 并且应用于实际项目中, 如油气勘探、水资源管理、气候变化研究等(段浩等, 2021; 龚仁彬等, 2021; Díaz, Vilches-Blázquez, 2022; Wu *et al.*, 2022a; Tang *et al.*, 2023)。

综上所述,知识图谱在地球科学领域已经得到了广泛关注和应用。然而,在构建这些知识图谱时,面临着一些挑战,其中之一就是地质学作为一个复杂而多样化的领域,其领域知识的表示和整合具有一定的复杂性。地质学是由多样化学科组成的,各种专业人员使用的概念的含义往往是根据他们当时的背景和工作习惯所做出的对于某一领域的特定看法,对于没有完全掌握的人来说,地质领域的定义是模糊的,对于同一个词,具有不同专业技能的地质学家有时会赋予不同的含义。现如今地质学知识广泛应用于石油勘探和生产、自然灾害评估和环境问题等项目中,很大程度上依赖于软件应用,专业的地质学家可以根据上下文来进行辨别,但是计算机尚无法实现。

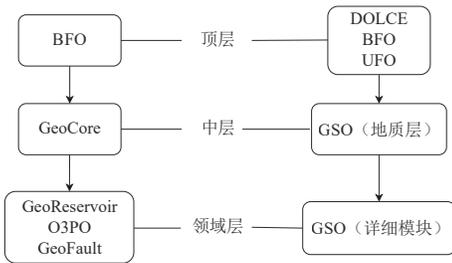


图 6 GeoCore 和 GSO 的体系结构对比
Fig. 6 Comparison of the architectures of GeoCore and GSO

为解决这些困难, Garcia 等(2020)提出了一个专门用于描述科学领域的地质学核心本体(GeoCore Ontology), Brodaric 和 Richard (2021)也构建了一个用来系统化表示地学关键知识的地质科学本体(GeoScience Ontology, 缩写为 GSO), 两者的体系结构如图 6 所示。GeoCore 相当于 GSO 的地质层, 同时也作为中间层延伸扩展成为了其它本体, 例如 GeoReservoir (Ciconeto *et al.*, 2022)、O3PO (Santos *et al.*, 2024) 和 GeoFault (Qu *et al.*, 2024) 等。

地质学作为地球科学的核心领域之一,其本体模型涵盖了地球内部和表层的各个方面,具有广泛的代表性和普适性。此外,地质学本体模型的构建方法已经相对成熟,可以为地球科学知识图谱的发展提供重要支持和启示。接下来将详细介绍上述两个地质学本体的构建方法。

2.1 地质本体模型: GeoCore 本体和 GSO 本体

GeoCore 本体是在 BFO 顶层本体的基础上提出的一个专门描述地学领域的核心本体,包含了地学领域内一组通用概念的严格定义。在 GeoCore 构建过程中, Garcia 等(2020)在地质学方面确定了有限数量的被所有地质学家广泛采用的概念,在本体论的研究中,参考了 Sure 等(2009)的本体工程方法和工具,分析基于 Guarino 和 Welty (2022)的 OntoClean 方法论。

GeoCore 本体中数量有限的概念包括地质对象、构成地质对象的物质、地质对象的边

界、地质对象的内部组织、地质对象的联系、地质对象的形成过程以及地质对象形成的地质时间间隔。对于每个概念，提出了一个同时考虑到实体的地质学和本体论性质的定义。按照 Abel 等(2016)的方法，通过概念之间的包含关系构建分类法，并将其与 BFO 顶层本体对齐。此外，还添加了额外概念补充这一框架，描述正式定义最初选定的概念所需的性质和关系，如图 7 所示。

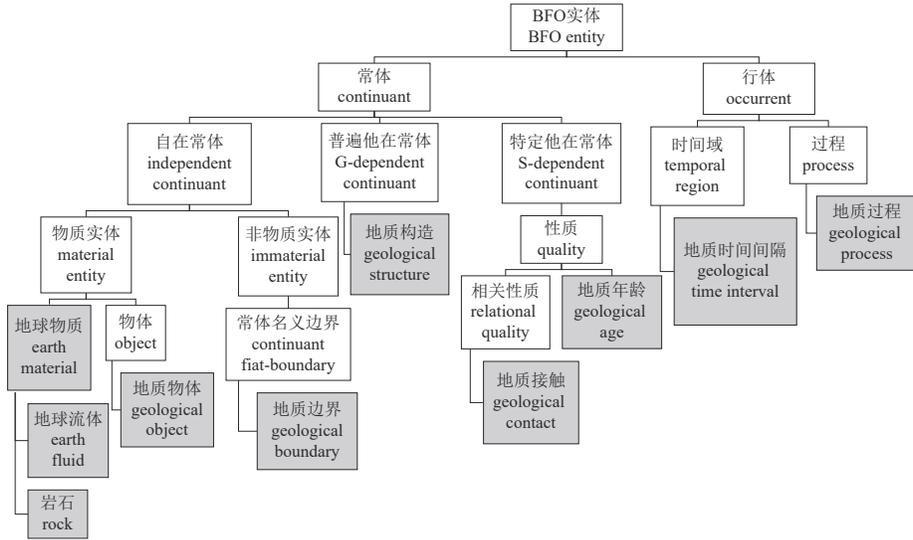


图 7 地质学核心本体概述(修改自 Garcia 等, 2020)

Fig. 7 Overview of the GeoCore Ontology (modified from Garcia et al, 2020)

GeoCore 本体可以整合各种要素，包括从微观到大陆尺度的任何大小的地质体，并可以整合描述这些要素的现有知识模型。此外，它还可以作为一个指南，用于在地质学领域中创建新的本体论，尤其是尚未形式化的子领域。GeoCore 本体为推导更专业的地质概念提供了基础，构成了整合不同领域本体的基线，也为消除常用地质解释概念存在的误解提供了解决方案，有利于将信息传达给外部地质用户以及在领域应用中实现集成。

地质科学本体(GeoScience Ontology, 缩写为 GSO)是对关键地质科学知识进行系统表述的框架，由三个层次组成：适用于任何学科的基础层，构成地质学任何方面根源的地质层，以及可以根据特定目的进行完善或补充的详细模块。

GSO 的顶层基础层在现有基础本体论的基础上进行了重新调整和整合，主要受到 DOLCE, BFO 和 UFO 的启发。语言和认知工程的描述性本体论(descriptive ontology for linguistic and cognitive engineering, 缩写为 DOLCE)是一个基于形式本体论的描述性本体，它以一种准确和可靠的方式为语言和认知工程提供共享的语义基础，通过定义和描述事物之间的关系和属性，使得计算机能够更好地理解和处理自然语言和认知概念。统一基础本体(unified foundational ontology, 缩写为 UFO)是近二十年来在哲学、认知科学、语言学和哲学逻辑等领域基于形式本体论整合发展起来的，包含了多组关于基本概念建模的微观理论，包括实体类型和关系类型(Guizzardi et al, 2022)。

在中间层，GSO 以北美地质图数据模型(north american geologic map data model, 缩写为

NADM)和 GeoSciML 为基础,在概念上进行延伸,形成了地质层,该层旨在全面表示地质学的各个方面,包括地质对象、材料、结构、环境、性质、作用、过程、事件、地质时间和地质关系等实体。

最后一层是延伸中间层的地学模块,例如各种地质构造、特定的时间尺度或各类岩石材料,这种模块化的方法可根据具体要求进行模块的替换或添加。

GSO 主要用途是用于三维地质建模的知识管理,可以被部署在没有互联网的地区,并与 3D 建模软件耦合。它是一个独立的产品,不会导入其它本体。GSO 采用 UML 表示时,使用 Sparx Enterprise Architect 工具,采用 OWL 表示,则使用原始文本编辑和 TopQuadrant TopBraid Composer、Protégé 等工具的组合。

前两个 GSO 层被序列化为不同的文件,“GSO-Common.ttl”为非基础地质层,“GSO-Geology.ttl”为地质层,构成第三个 GSO 层的模块也是不同的 ttl 文件,每个地质方面一个。GSO Common 模块包含最一般的非地质实体。GSO 地质层(GSO-Common.ttl)包含最上层的地质实体,每个地质实体专门化一个共同的实体,或者是持久体(Endurant)、持续体(Perdurant)、特征(Feature)、情境(Situation)或某个子类型,如图 8 所示。

GeoCore 本体相当于 GSO 的中间层,可以对 GeoCore 本体进行进一步的扩展,比如建立岩石、地层学或者构造地质学的领域本体,但是 Garcia 等(2020)并未进行实例检验,仅提出了想法。而在 GSO 中,已经构建了详细模块,每个 GSO 模块至少需要 GSO-Common,地质模块额外需要 GSO-Geology 和少量其它模块,应用程序可只部署必要的模块。

2.2 其它本体模型

地质数据与知识图谱的融合已被证明是有效的,并且可增强这些数据集之间的相互连接(Abu-Salih, 2021)。除了上述提到的两个地质本体以外,还存在其它模型在地质学领域的应用,例如:Zhang 等(2013)提出了 GeoDeepDive 模型,该模型可以帮助地质学家从地质期刊论文的文本、表格和图形中发现隐藏的信息和知识;PaleoDeepDive (Peters *et al*, 2014)可以从出版物中的异构文本、表格和图形中自动定位和提取数据,它可以很容易地容纳复杂的数据类型,产生了一个概率数据库,随着信息的增加而系统改进;Zhu 等(2017)在地质数据深度挖掘智能系统展示了知识图谱的应用,利用 Baike.com 和当地地质文件构建了知识图谱。

在地质文献挖掘领域,前人作了大量研究。地质学家在参与地质调查和科学研究时,常常使用地质报告来记录调查结果和成果。尽管有如此丰富的数据来源,仍有大量知识等待挖掘和分析。Wang 等(2018)使用自然语言处理(natural language processing, 缩写为 NLP)方法和知识图谱从地质文献中提取和可视化知识图谱,为重用大量的非结构化文献提供了一个新的视角。Lv 等(2022)提出了 BERT-BiGRU-CRF,这是一种基于深度学习的地质命名实体识别模型,专门针对语言不规则性而设计。Wang 等(2022)提出了一种通过三元组抽取从地质报告中自动抽取信息的方法,然后从抽取的实体和关系中自动构建地质知识图谱。

知识图谱也可以用来辅助寻找矿产,比如建立找矿模型(Liu *et al*, 2022)和从未充分利用的勘探报告中自动提取与矿化和成矿条件相关的地质信息(Enkhsaikhan *et al*, 2021),并且已经应用于具体的某一类矿产类型(周永章等, 2021; 冉一早等, 2023; 张春菊等, 2023)。

对于油气勘探领域,有些学者将知识图谱用于油气层识别(刘国强等, 2022)和油气管道风险评估(吴张中, 2023)。Guichet 等(2019)创建了一个系统,能够在大型文档储库中准确定位与回答石油勘探相关的具体问题,特别是与烃源岩表征相关的问题,采用机器学习系统

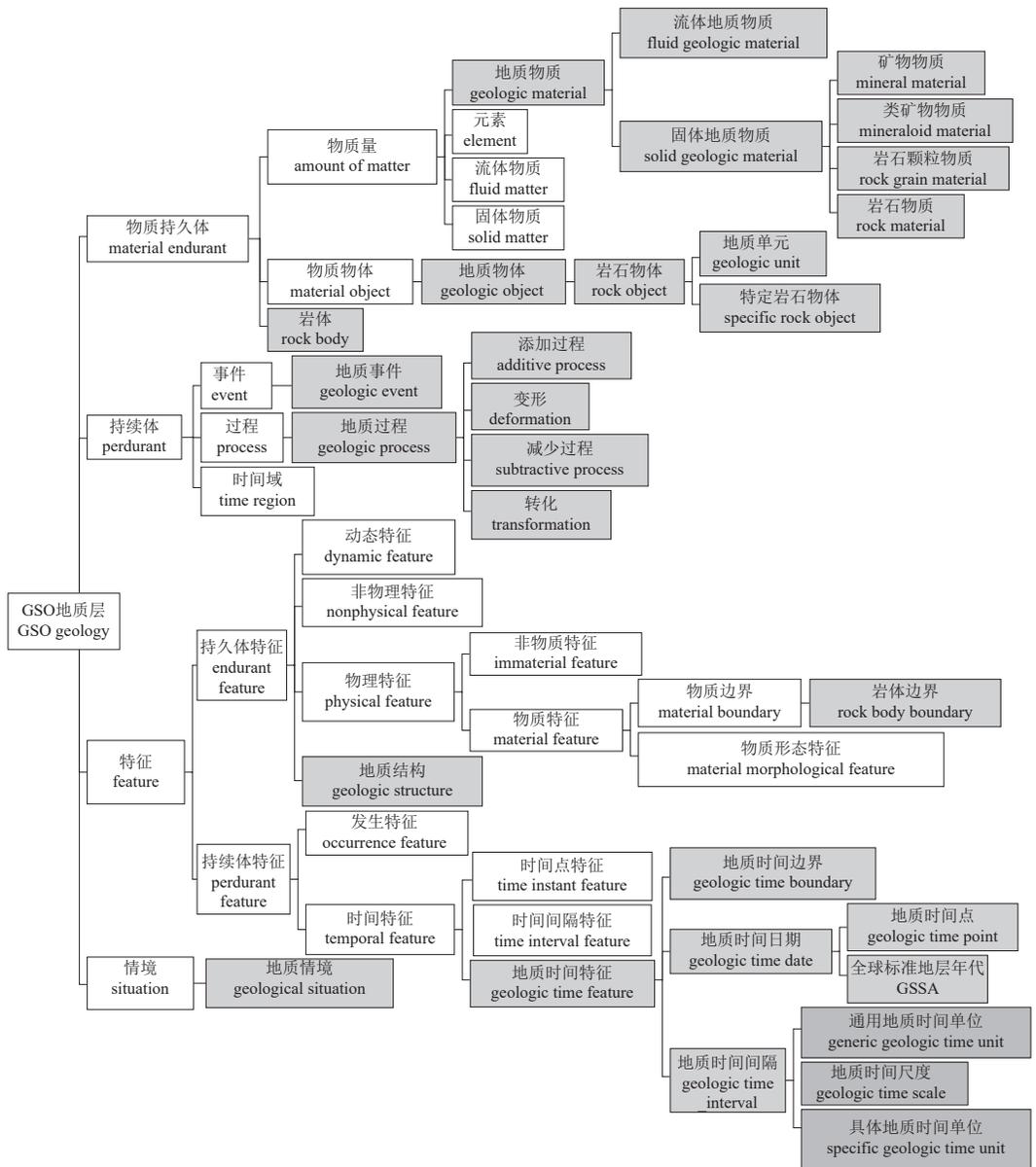


图 8 地质科学本体概述 (修改自 Brodaric 和 Richard, 2021)

Fig. 8 Overview of the GSO Ontology (modified from Brodaric and Richard, 2021)

WATSON (IBM) 来支持地质学家在区域地质研究中的应用, 旨在利用机器学习算法, 提高从海量非结构化科学文档中收集相关地球科学家数据的效率。

在自然灾害领域, 知识图谱也得到了广泛应用, 常见的就是应用于地质灾害、台风灾害、森林火灾和洪涝灾害 (Mughal et al., 2021; Ge et al., 2022a; Zhang, Wang, 2022; 邱芹军等, 2023; 王益鹏等, 2023) 等。除了实例应用之外, 也有学者利用知识图谱对文献进行分析, 获取现今灾害风险感知的主要研究对象 (Yu et al., 2021), 或者与遥感信息、相关地理信息等相结合, 建立灾害预测模型 (Ge et al., 2022b) 和应急响应知识概念模型 (Liu et al., 2021),

有效降低灾害损失。

也有学者将知识图谱应用于遥感领域, Li 等(2021)提出一种基于深度对齐网络(Deep Alignment Network, 缩写为 DAN)的方法, 通过跨模态匹配实现了遥感图像分类中的零样本学习, 取得了比其它方法更好的性能。Hao 等(2021)根据测绘遥感大数据的异构、分散和动态更新特征, 构建了面向测绘遥感应用的知识图谱。Aldana-Martin 等(2022)提出了一个语义模型, 用于部分满足对地观测(EO)卫星遥感产品的数据(和元数据)合并、集成、推理和链接的需求。Sun 等(2022)讨论了遥感图像解译的主要挑战, 并系统地综述了基于语义图的遥感知识表示和理解的典型方法, 包括本体模型、地学信息图谱和语义知识图谱。

在气候学领域, Rising 等(2022)对气候风险进行了概述, 并根据这些风险缺乏稳健评估的原因进行了本体分析, 还提出了一种方法, 以综合评估或猜测这些风险时考虑相互依赖、避免重复计算和明确假设。Wu 等(2022b)提出了一种能够自动集成远程气候知识图谱数据和本地表格数据的分析流程工作, 以提高特定气候分析任务的数据可用性。此外他们还创建了一个可互操作的气候数据知识图谱平台 LinkClimate, 这个图谱将多个气候数据和其它数据源集成到一个服务中, 并利用 Web 技术(例如 HTTP)进行多源气候数据分析(Wu *et al.*, 2022a)。

此外, 知识图谱也可以应用于构建数字高程模型(DEM)、水环境智能监测、表征水质、海上风电场检测、高山滑雪风险评估、地铁车站地震风险综合评估和地震事件检测(Wang *et al.*, 2021b; Zhao *et al.*, 2021; Díaz, Vilches-Blázquez, 2022; Falanga *et al.*, 2022; Hoerer, Kuenzer, 2022; Mezni *et al.*, 2022; Cui *et al.*, 2023)等领域。然而, 地理知识图谱主要描述静态事实, 难以表征变化, 从而严重限制了其在地理时空过程中的应用, 为了解决这个问题, Zheng 等(2022)提出了一种称为地理演化知识图谱(Geographic Evolutionary Knowledge Graph, 缩写为 GEKG)的模型, 该模型将地理要素的时空特征和演化规律纳入分析, 以实现更好的地理知识表征。

总之, 地球科学知识图谱应用领域广泛, 包括地质学、气候学、地理学等多个学科领域(表 2), 为科学研究、决策支持和可持续发展提供了重要的数据和知识基础。

3 地学知识图谱项目

3.1 深时数字地球计划(DDE)

人类很早就开始研究宇宙演化、地球演化和生命演化的科学问题, 地球历史中保存着信息丰富但不完整的地质记录, 研究地球的深时历史有助于揭示演化机制、气候变化、自然资源 and 地球的未来。演绎推理和归纳推理已被广泛用于研究地球历史, 但对于不容易通过上述两种方法发现的模式, 溯因推理是一种有效方法, 它源于对大量可靠数据的积累和分析, 与前提或概括无关。随着海量地球数据的积累, 地球科学家计划通过溯因发现来改进地球科学研究, 地球深时数据以不同的格式存储在不同的文献和数据库中。为了整合和协调这些数据, 已建立了多个专题数据设施, 如 Paleobiology Database (PBDB)、Macrostrat、EarthChem 和 RRUFF 等, 表 3 所示为目前正在使用或正在开发的 DDE 数据库和门户, GBDB 和 OneStratigraphy 是 DDE 正在开发的数据库, 其余为 DDE 使用的独立数据系统。当利用深度时间数据库进行溯因时, 会遇到诸多问题, 但大数据和人工智能为解决这些问题带来了机遇。国际地质科学联合会(International Union of Geological Sciences, 缩写为 IUGS)发起了深时数字地球计划(DDE), 并与全球各国的地质调查机构、专业协会、学术机构和科学家开展合作

表 2 其它本体模型概览

Table 2 Overview of other ontology models

领域	本体模型
地质学	地质期刊论文信息提取(GeoDeepDive)(Zhang et al, 2013) 合成古生物数据库的机器读取系统(PaleoDeepDive)(Peters et al, 2014) 面向地质数据的知识图谱智能学习(Zhu et al, 2017)
地质文献挖掘	地学文献信息抽取(Wang et al, 2018) 中文命名实体识别(BERT-BiGRU-CRF)(Lv et al, 2022) 地质报告理解(GeoERE-Net)(Wang et al, 2022)
矿产	成矿条件(Enkhsaikhan et al, 2021) 钦杭成矿带斑岩铜矿知识图谱(周永章等, 2021) 找矿概念模型(Liu et al, 2022) 铌钽矿床知识图谱(冉一早等, 2023) 金矿知识图谱(张春菊等, 2023)
油气勘探	油气盆地勘探地学文献信息提取(Guichet et al, 2019) 神经网络油气层评价模型(KPNFE)(刘国强等, 2022) 油气管道地质灾害风险管理(吴张中, 2023)
自然灾害	河流流量与防洪减灾模型(ORFFM)(Mughal et al, 2021) 灾害风险感知的科学计量分析(Yu et al, 2021) 地质灾害应急响应(Liu et al, 2021) 林火预测(Ge et al, 2022b) 全球洪水灾害研究图表分析(Zhang, Wang, 2022) 灾害预测知识图谱(Ge et al, 2022a) 地质灾害链知识图谱(邱芹军等, 2023) 台风灾害事件知识图谱(王益鹏等, 2023)
遥感	遥感图像场景分类(Li et al, 2021) 测绘遥感应用(Hao et al, 2021) 对地观测遥感语义建模(Aldana-Martín et al, 2022) 遥感图像解译方法综述(Sun et al, 2022)
气候学	气候变化的缺失风险(Rising et al, 2022) 气候数据知识图谱平台(LinkClimate)(Wu et al, 2022a) 气候分析(Wu et al, 2022b)
其它	数字高程模型构建(Zhao et al, 2021) 高山滑雪赛事风险评估(Wang et al, 2021b) 水环境智能检测(SmartWater)(Mezni et al, 2022) 表征水质数据集(Díaz, Vilches-Blázquez, 2022) 海上风电场检测(SyntEO)(Hoerer, Kuenzer, 2022) 地震事件检测(Falanga et al, 2022) 地理时空进程的知识表示模型(Zheng et al, 2022) 地铁站综合地震风险评估(Cui et al, 2023)

(Wang et al, 2021a).

DDE 的目标是链接和协调全球深层地球数据, 共享全球地学知识, 旨在通过深层时间刺激数据驱动的发展来研究地球的演化. 地球演化涉及四个主题: 地球上的生命、地球物质、地理和气候. 为实现其使命和愿景, 计划包含三个主要组成成分: 计划管理委员会、卓越中心和工作、平台以及任务组. DDE 将基于现有的地球深时知识体系, 进行开放平台的开发. 研究计划的第一步基于现有的地球深时知识体系建立模型, 第二步是构建可互相操作的地球深时数据基础设施, 第三步是开发深时地球开放平台. 程序的执行分为四个阶段: 第一阶段建立一个组织结构; 第二阶段形成初步团队, 构建深时地球数据库; 第三阶段开发算法和技术; 第四阶段解决综合性科学问题.

目前, 在地球科学领域, 知识图谱的构建采用了各种语言和工具, 但不同知识图谱之间的关系通常不明确, 这给知识图谱的整合和重复使用带来了挑战. 考虑到目前地学领域知识

表 3 DDE 数据库和门户(修改自 Wang 等, 2021a)
Table 3 DDE database and portal (modified from Wang *et al.*, 2021a)

数据库或门户网站	领域	网址
EarthChem	获取全球地球化学和岩石学数据综合(PetDB, EarthChem Portal, LEPR, traceDs); EarthChem Library将地球化学、岩石学和矿物学数据作为出版商推荐的可信存储库进行发布和归档	https://www.earthchem.org
Geobiodiversity Database (GBDB)	分段式地层和古生物信息管理与分析集成系统	http://geobiodiversity.com/home
GeoDeepDive and PaleoDeepDive	数字图书馆和网络基础设施有助于发现和利用已发表文献中的地质数据和知识	https://geodeepdive.org
Macrostrat	用于与沉积岩、火成岩和变质岩的空间和时间分布以及从中提取的数据相关的地质数据的聚合和分布的协作平台	https://macrostrat.org
Mindat	世界上最大的矿物、岩石、陨石及其产地的开放数据库	https://zh.mindat.org
OneGeology Portal	全球范围内地质图数据和相关地学数据的比例尺 $\geq 1:100$ 万	https://portal.onegeology.org/OnegeologyGlobal
OneStratigraphy Database	平台设计用于共享和使用地层数据, 包括地层数据的集成、管理、可视化和分析	https://onestratigraphy.ddeworld.org
Paleobiology Database (PBDB)	全球范围内, 基于收集的所有地质年代的生物的发生和分类数据, 以及数据服务, 允许轻松获取数据独立开发分析工具, 可视化软件和应用程序	https://paleobiodb.org
PANGAEA	地球与环境科学数据发布者; 开放存取图书馆旨在归档、发布和分发来自地球系统研究的地理参考数据	https://pangaea.de
RRUFF	高质量的光谱数据(包括X射线衍射、拉曼光谱和电子探针分析)来自表征良好的矿物	https://rruff.info
PALEOMAP	在过去的1100百万年中, 板块构造重建	http://scotese.com
GPlates	这是一个开放源代码的软件, 提供了交互式板块构造重建、地理信息系统功能和栅格数据可视化的新组合	https://www.gplates.org

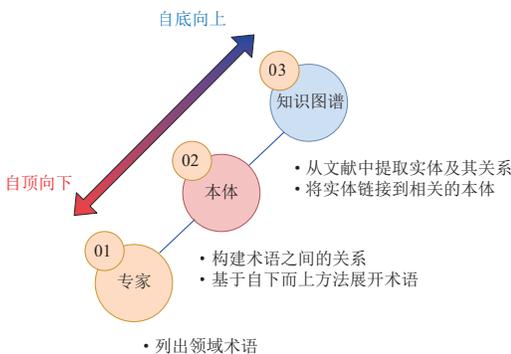


图 9 构建 DDE 知识图谱的工作流程
(修改自 Hu 等, 2023)

Fig. 9 Construction workflow for DDE knowledge graph (modified from Hu *et al.*, 2023)

图谱构建的现状, DDE 项目旨在建立统一的表示模型, 以构建地学领域的综合知识图谱。为了实现这一目标, DDE 知识图谱遵循三阶段构建方案(图 9)。第一阶段的重点是开发知识系统, 第二阶段建立术语之间的关系和属性, 从而形成本体。DDE 的本体分为两类: 基础本体, 包括空间本体和地质时间本体等地球科学领域常用的本体; 特定领域本体, 包括来自地球科学各分支学科的专门内容。目前, 在 DDE 框架下已经初步构建了 20 个特定领域的本体, 包含 61 937 个节点和 62 610 个关系(表 4)。为实现对专家构建的本体的开放访问, DDE 开发了地学知识图谱协同编辑器。在四年的时间里(2019—2023), DDE 知识图谱在该平台上实现了第一阶段的建设目标, 现已进入第二阶段的本体建设阶段(Hu *et al.*, 2023)。

DDE 计划面临多种挑战, 需要吸引更多的科学家和科学社群参与, 整合各种语言和格式的数据, 建设全球数据基础设施, 同时要推进国际合作, 连接不同数据和网络基础设施, 并协商可持续发展。

总之, 通过整合地球深部数据、地学知识以及数据科学和人工智能技术, DDE 解决了地

球演化研究中的重大挑战,促进了数据的整合和研究,开创了地球科学中跨学科溯源发现的新时代。DDE 将创造一个开放访问的数据资源,整合了地球叙述的过去、现在和未来的方方面面,这将带来全新的视角,改变我们对地球和宇宙的认识。

3.2 其它项目

DDE 致力于构建综合的地学知识图谱,通过整合各种地学领域的知识和数据资源,以支持地球科学的研究和应用。除了 DDE 项目,国内外还有许多其它组织和计划也在推动地球科学知识图谱的发展,各项的侧重点各有不同。

国际地质科学联合会地理信息委员会(The Commission for Geoinformation within the International Union of Geological Sciences, 缩写为 IUGS-CGI)是标准化地球科学词汇和地质数据图式的促进者。IUGS-CGI 的一部分成果已被整合到 OneGeology、OneGeology-Europe 和 INSPIRE 计划中,以协调来自分布式来源的地质数据(Ma, 2022)。

OneGeology 项目首要目标是提高地质图数据的可获取性,使所有国家都能参与交流专门知识和技能并加速地学领域的互操作性和新标准(GeoSciML)(Jackson, 2008)。通过整合来自不同来源的地质数据,提供了高分辨率的地质图数据,使得地质数据在全球范围内更易于获取和使用。

OneGeology-Europe 项目是 OneGeology 项目的一部分,旨在整合和共享欧洲地质地球科学数据和信息,为科学家、工程师和公众提供更全面、更便捷的地质地球科学数据资源。OneGeology-Europe 项目展示了知识图谱在地质图集成和服务中的智能化应用(Ma, 2022)。GeoSciML 和词汇表被应用于 OneGeology-Europe 项目中,该项目使用了来自 20 个不同数据供应商的分布式 Web 服务绘制了 1:100 万欧洲地质图(Laxton, 2017)。

从 OneGeology-Europe 项目中得到的经验教训促进了 INSPIRE 地质数据规范的发展,INSPIRE 计划旨在促进欧洲范围内的空间数据和地理信息的互操作性和共享,建立统一的空间信息基础设施,支持环境管理、城市规划等领域的决策和行动。GeoSciML 数据模型是一种标准化的数据模型,用于描述地球科学领域的地质地球科学数据,促进数据的交换、共享和集成,提高地球科学研究和应用的效率和准确性。GeoSciML 可以扩展 INSPIRE 数据模型,为地质图融合提供基础(Laxton, 2017)。

EarthCube 项目主要目标是建立学科之间的连接,促进现代地球科学的发展和变革。它是一个由美国国家科学基金会支持的社区驱动项目,超过 2500 名地球、大气、海洋、计算机、信息和社会科学家,以及教育学家和数据管理者等人员参与了对 EarthCube 项目概念的阐述(Richard *et al*, 2014)。该项目在发展过程中也取得了一些进展,比如 Zhou 等(2020)提出了一个来自于 GeoLink 项目的真实数据集,作为潜在的复杂本体对齐基准,该数据集包括 GeoLink Base Ontology (GBO) 和 GeoLink Modular Ontology (GMO) 两个本体,以及来自不同机构的领域专家共同开发的手动创建的参考对齐。GeoLink 通过 SPARQL 协议和 RDF 查询语言

表 4 DDE 领域本体,修改自 Hu 等(2023)

Table 4 DDE domain ontology, modified from Hu *et al* (2023)

领域	节点	领域	节点
古生物学	24936	数学地球科学	765
地层学	1268	地磁古地磁学	2434
沉积学	2675	工程地质学	3446
古地理学	2831	石油地质学	2390
矿物学	5665	地热学	907
火成岩岩石学	1670	水文地质学	806
变质岩石学	1028	地貌学	1958
矿床	605	表生地球化学	3690
构造地质学与 大地构造学	1291	地质填图	1990
		地质年代学	424

(SPARAL)终端对外公开,包含超过4500万个RDF三元组,以及一组本体和地理可视化工具(Cheatham *et al.*, 2018). GeoLink旨在通过使用本体来提高7个地学数据存储库的数据检索、重用和集成(Krisnadhi *et al.*, 2015).

LinkedGeoData项目是一个基于开放地理数据的开源项目,旨在提供一个集成了地理和语义信息的全球地理数据库.它创建了一个全球覆盖的大型知识图谱,其数据源包括OpenStreetMap(OSM)和其它数据源,并且这些数据源之间相互链接.通过LinkedGeoData,用户可以访问和查询各种地理信息,例如地理实体(如城市、国家、地标等)、地理关系和属性等.Ding等(2021)报告了使用虚拟知识图谱(Virtual Knowledge Graph,缩写为VKG)技术将LinkedGeoData中的关系数据库作为SPARQL端点持续努力的情况.

此外,NASA建立了SWEET本体,旨在通过软件理解Web资源的语义,提高地球科学数据的发现和使用,涵盖了地球和环境科学的200多个学科领域和6000多个概念(Raskin, Pan, 2005; Ma, 2022).

4 讨论和展望

4.1 面临的困难和挑战

知识图谱面临着多个挑战,包括可扩展性、质量、多样性、动态性和可用性,这些挑战不太可能被完全解决,但随着技术和工具的发展,它们将逐渐得到解决(Hogan *et al.*, 2021). Roldán-Molina等(2021)提出了一种评估本体质量的方法(定量和图形化),并解决了本体不一致性的问题,以最大限度地减少设计缺陷.此外,Chen等(2020)也详细介绍了各种知识图谱补全(knowledge graph completion,缩写为KGC)技术的特点、优势和适用领域.

上述挑战不仅存在于地球科学领域,也是所有知识图谱构建中普遍存在的.然而,由于地球科学的复杂性和多样性,地球科学中的知识图谱构建面临着一些独特的困难.Ma(2022)在论文中讨论了地球科学知识图谱构建和应用中的几个重要问题,包括KG实体消歧和识别、质量度量、语义定义和推理能力、KG的演化与版本、KG之间的互联与扩容,以及安全、隐私和道德问题.这些问题对于KG的有效使用和发展至关重要.除此之外,对于特定领域的知识图谱所面临的局限和不足,Abu-Salih(2021)也进行了总结和概括,主要包括:① KG数据质量、隐私性和可信性;② 知识资源与语义扩展;③ KG构造算法;④ 时间感知的KGs;⑤ KG评价;⑥ Big KGs中的计算性能;⑦ 特定领域的KG推理;⑧ 领域特定KGs的可用性;⑨ 领域特异性KGs的进一步研究空间.

4.2 地球科学知识图谱应用展望

地球科学中的KG创建和应用的未来工作具有广阔的发展空间和灵活性:首先,地球科学知识图谱可进一步提高数据的可访问性和可利用性,将大量多源异构数据整合到知识图谱中,并实现数据的互操作性,使得不同数据之间可以无缝连接和交互;其次,地球科学知识图谱需要更加精确和丰富的知识表示方式,包括地质、气候、海洋、生态等多个领域的知识,并具备推理能力,能够自动地推导和发现新的知识,从而加速科学研究的进展.此外,地球科学知识图谱的发展需要跨学科的融合和合作,将地球科学与计算机科学、人工智能、数据科学等领域的知识相结合,共同推动知识图谱的构建和应用,促进不同学科之间的交流与合作;最后,为了更好地展示和利用地球科学知识图谱,未来需要开发出直观、易用的可视化和交互界面,使科学家、决策者和公众能够方便地浏览、查询和探索知识图谱中的信

息,从而更好地理解地球系统和相关问题。

在地球科学领域,大型语言模型(large language models, 缩写为 LLMs)的引入为知识图谱构建和应用带来了新的机遇。Deng 等(2024)将 LLMs 引入地球科学领域,首次提出了地球科学领域的第一个 LLM 模型 K2,并开发了一套资源来促进地球科学领域内的 LLMs 研究。大型语言模型可以通过学习大规模文本数据来理解和生成自然语言,但它并不总是具有深度的领域特定知识,知识图谱提供了结构化的、领域特定的知识,弥补了大型语言模型在特定领域知识方面的不足。将大型语言模型和知识图谱结合使用可以提高自然语言处理的性能,促进对地球科学领域中复杂问题更深入的理解和分析。

地震科学作为地球科学的重要分支,涉及到地质学、地球物理学、工程地震学等多个一级学科的交叉。然而,知识图谱在地震科学领域的应用目前还存在着较大的空白,迫切需要在现有的模型基础上开展深入研究。地震科学知识图谱可以从上文提到的 BFO 本体进行延伸,利用其通用性,将地震科学中的基本概念与 BFO 中的一般概念进行对应,然后通过扩展添加地震科学特定的概念。确定地震科学概念之间的关系,比如地震事件与地质结构、地震波的传播路径等。整合地震领域的数据,与专业人士合作,确保本体的设计和知识图谱的建立符合实际研究需求。地震科学知识图谱的建立有助于整合多领域信息,推动地震科学领域的发展。

总体而言,地球科学知识图谱的发展将是一个不断演进和完善的过程,通过持续的技术创新和学科交叉,将为地球科学研究、决策制定和公众教育等领域带来更多的机遇和益处。

5 总结

地球科学知识图谱是整合多源地球科学数据的工具,有助于科学家理解地球系统的复杂性,GeoCore 和 GSO 地质本体的建成证明地质数据与知识图谱结合良好,此外,它还广泛应用于地学文献挖掘、气候、自然灾害和遥感等领域。DDE 计划、OneGeology-Europe 项目和 INSPIRE 计划等推动了数据集成和共享,促进了地球科学知识图谱的发展。然而,虽然有学者已经在质量评价和知识图谱补充方面进行了研究,但还是面临可扩展性、质量、多样性、动态性和可用性等方面的挑战。在未来的发展中,依旧需要技术完善和国际合作的支持,并且可以与大型语言模型等先进技术相结合,以促进地球科学知识图谱的发展。此外,在地震科学领域,知识图谱通过整合和智能分析地震相关数据与知识,为地震研究人员提供更准确的灾害预测、风险评估以及快速响应和决策支持的能力。

国家超算成都中心郑亮博士和阿里云数字政府首席架构师魏文麟在调研国内地球大数据基础设施、算力和平台技术发展情况的过程中分享了前沿研究成果、提供了专业建议和指导并协助数据收集和分析,审稿人对本文提出了建设性的意见和建议,作者在此一并表示感谢。

参 考 文 献

- 杜志强,李钰,张叶廷,谭玉琪,赵文豪. 2020. 自然灾害应急知识图谱构建方法研究[J]. 武汉大学学报:信息科学版, 45(9): 1344-1355.
- Du Z Q, Li Y, Zhang Y T, Tan Y Q, Zhao W H. 2020. Knowledge graph construction method on natural disaster emergency[J].

Geomatics and Information Science of Wuhan University, **45**(9): 1344–1355 (in Chinese).

- 段浩, 韩昆, 赵红莉, 蒋云钟, 李豪, 毛文山. 2021. 水利综合知识图谱构建研究[J]. 水利学报, **52**(8): 948–958.
- Duan H, Han K, Zhao H L, Jiang Y Z, Li H, Mao W S. 2021. Research on water conservancy comprehensive knowledge graph construction[J]. *Journal of Hydraulic Engineering*, **52**(8): 948–958 (in Chinese).
- 龚仁彬, 杨燕子, 任义丽, 张晓宇. 2021. 知识图谱在石油勘探开发领域的应用现状及发展趋势[J]. 信息系统工程, (9): 16–18.
- Gong R B, Yang Y Z, Ren Y L, Zhang X Y. 2021. Application status and development trend of knowledge map in petroleum exploration and development[J]. *China CIO News*, (9): 16–18 (in Chinese).
- 刘国强, 龚仁彬, 石玉江, 王珍珍, 米兰, 袁超, 钟吉彬. 2022. 油气层测井知识图谱构建及其智能识别方法[J]. *石油勘探与开发*, **49**(3): 502–512.
- Liu G Q, Gong R B, Shi Y J, Wang Z Z, Mi L, Yuan C, Zhong J B. 2022. Construction of well logging knowledge graph and intelligent identification method of hydrocarbon-bearing formation[J]. *Petroleum Exploration and Development*, **49**(3): 502–512 (in Chinese).
- 刘娇, 李杨, 段宏, 刘瑶, 秦志光. 2016. 知识图谱构建技术综述[J]. *计算机研究与发展*, **53**(3): 582–600.
- Liu Q, Li Y, Duan H, Liu Y, Qin Z G. 2016. Knowledge graph construction techniques[J]. *Journal of Computer Research and Development*, **53**(3): 582–600 (in Chinese).
- 毛文山, 赵红莉, 蒋云钟, 段浩, 郝震. 2019. 基于文献计量学的国内水生态环境研究知识图谱构建与应用[J]. 水利学报, **50**(11): 1400–1416.
- Mao W S, Zhao H L, Jiang Y Z, Duan H, Hao Z. 2019. Construction and application of knowledge graph of domestic water eco-environment based on bibliometrics[J]. *Journal of Hydraulic Engineering*, **50**(11): 1400–1416 (in Chinese).
- 齐浩, 董少春, 张丽丽, 胡欢, 樊隽轩. 2020. 地球科学知识图谱的构建与展望[J]. 高校地质学报, **26**(1): 2–10.
- Qi H, Dong S C, Zhang L L, Hu H, Fan J X. 2020. Construction of Earth science knowledge graph and its future perspectives[J]. *Geological Journal of China Universities*, **26**(1): 2–10 (in Chinese).
- 邱芹军, 吴亮, 马凯, 谢忠, 陶留锋. 2023. 面向灾害应急响应的地质灾害链知识图谱构建方法[J]. 地球科学, **48**(5): 1875–1891.
- Qiu Q J, Wu L, Ma K, Xie Z, Tao L F. 2023. A knowledge graph construction method for geohazard chain for disaster emergency response[J]. *Earth Science*, **48**(5): 1875–1891 (in Chinese).
- 冉一早, 董少春, 王汝成, 胡欢, 方思源, 常丰年, 孔嘉. 2023. 铌钽矿床知识图谱的构建及应用实践[J]. 高校地质学报, **29**(3): 359–371.
- Ran Y Z, Dong S C, Wang R C, Hu H, Fang S Y, Chang F N, Kong J. 2023. Construction of Nb-Ta deposit knowledge graph and its application[J]. *Geological Journal of China Universities*, **29**(3): 359–371 (in Chinese).
- 王益鹏, 张雪英, 党玉龙, 叶鹏. 2023. 顾及时空过程的台风灾害事件知识图谱表示方法[J]. 地球信息科学学报, **25**(6): 1228–1239.
- Wang Y P, Zhang X Y, Dang Y L, Ye P. 2023. Knowledge graph representation of typhoon disaster events based on spatiotemporal processes[J]. *Journal of Geo-Information Science*, **25**(6): 1228–1239 (in Chinese).
- 吴张中. 2023. 油气管道地质灾害风险管理知识图谱构建与应用[J]. *油气储运*, **42**(3): 241–248.
- Wu Z Z. 2023. Construction and application of knowledge graph of geohazard risk management for oil and gas pipeline[J]. *Oil & Gas Storage and Transportation*, **42**(3): 241–248 (in Chinese).
- 张春菊, 刘文聪, 张雪英, 叶鹏, 汪陈, 朱少楠, 张达玉. 2023. 基于本体的金矿知识图谱构建方法[J]. 地球信息科学学报, **25**(7): 1269–1281.
- Zhang C J, Liu W C, Zhang X Y, Ye P, Wang C, Zhu S N, Zhang D Y. 2023. Knowledge graph construction method of gold mine based on ontology[J]. *Journal of Geo-Information Science*, **25**(7): 1269–1281 (in Chinese).
- 张吉祥, 张祥森, 武长旭, 赵增顺. 2022. 知识图谱构建技术综述[J]. 计算机工程, **48**(3): 23–37.
- Zhang J X, Zhang X S, Wu C X, Zhao Z S. 2022. Survey of knowledge graph construction techniques[J]. *Computer Engineering*, **48**(3): 23–37 (in Chinese).
- 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. 钦杭成矿带斑岩铜矿知识图谱构建

及应用展望[J]. 地学前缘, **28**(3): 67–75.

- Zhou Y Z, Zhang Q L, Huang Y J, Yang W, Xiao F, Ji J J, Han F, Tang L, Ouyang C, Shen W J. 2021. Constructing knowledge graph for the porphyry copper deposit in the Qinzhou-Hangzhou bay area: Insight into knowledge graph based mineral resource prediction and evaluation[J]. *Earth Science Frontiers*, **28**(3): 67–75 (in Chinese).
- 朱彦, 郑捷, 李晓瑛, 杨啸林, 何勇群. 2021. 基本形式化本体及其中文版介绍[J]. 医学信息学杂志, **42**(1): 24–28.
- Zhu Y, Zheng J, Li X Y, Yang X L, He Y Q. 2021. Introduction to basic formal ontology and its Chinese version[J]. *Journal of Medical Informatics*, **42**(1): 24–28 (in Chinese).
- Abel M, Perrin M, Carbonera J L, Garcia L. 2016. Ontologies and data models: Essential properties and data modeling for petroleum exploration[J]. *Found J Prof Petrol Data Manag Assoc*, **3**(1): 18–19.
- Abu-Salih B. 2021. Domain-specific knowledge graphs: A survey[J]. *J Netw Comput Appl*, **185**: 103076.
- Aldana-Martín J F, García-Nieto J, Del Mar Roldán-García M, Aldana-Montes J F. 2022. Semantic modelling of earth observation remote sensing[J]. *Expert Syst Appl*, **187**: 115838.
- Bassett L. 2015. *Introduction to JavaScript Object Notation: A To-The-Point Guide to JSON*[M]. Sebastopol: O'Reilly Media, Inc: 1–4.
- Brodaric B, Richard S M. 2021. *The GeoScience Ontology Reference*[M]. Ottawa: Geological Survey of Canada: 34.
- Chaudhri V K, Baru C, Chittar N, Dong X L, Genesereth M, Hendler J, Kalyanpur A, Lenat D B, Sequeda J, Vrandečić D, Wang K S. 2022. Knowledge graphs: Introduction, history, and perspectives[J]. *AI Mag*, **43**(1): 17–29.
- Cheatham M, Krisnadhi A, Amini R, Hitzler P, Janowicz K, Shepherd A, Narock T, Jones M, Ji P. 2018. The GeoLink knowledge graph[J]. *Big Earth Data*, **2**(2): 131–143.
- Chen T, Wang X W, Yue T, W Bai X Y, Le C X, Wang W P. 2023. Enhancing abstractive summarization with extracted knowledge graphs and multi-source transformers[J]. *Appl Sci*, **13**(13): 7753.
- Chen X L, Xie H R, Li Z X, Cheng G. 2021. Topic analysis and development in knowledge graph research: A bibliometric review on three decades[J]. *Neurocomputing*, **461**: 497–515.
- Chen Z, Wang Y H, Zhao B, Cheng J, Zhao X, Duan Z T. 2020. Knowledge graph completion: A review[J]. *IEEE Access*, **8**: 192435–192456.
- Cicconeto F, Vieira L V, Abel M, Dos Santos Alvarenga R, Carbonera J L, Garcia L F. 2022. GeoReservoir: An ontology for deep-marine depositional system geometry description[J]. *Comput Geosci*, **159**: 105005.
- Cui C Y, Xu M Z, Xu C S, Zhang P, Zhao J T. 2023. An ontology-based probabilistic framework for comprehensive seismic risk evaluation of subway stations by combining Monte Carlo simulation[J]. *Tunn Undergr Sp Tech*, **135**: 105055.
- Deng C, Zhang T H, He Z M, Chen Q Y, Shi Y Y, Xu Y, Fu L Y, Zhang W N, Wang X B, Zhou C H, Lin Z H, He J X. 2024. K2: A foundation language model for geoscience knowledge understanding and utilization[C]//*Proceedings of the 17th ACM International Conference on Web Search and Data Mining*. New York: Association for computing Machinery: 161–170.
- Díaz J D R, Vilches-Blázquez L M. 2022. Characterizing water quality datasets through multi-dimensional knowledge graphs: A case study of the Bogota river basin[J]. *J Hydroinform*, **24**(2): 295–314.
- Ding L F, Xiao G H, Pano A, Stadler C, Calvanese D. 2021. Towards the next generation of the LinkedGeoData project using virtual knowledge graphs[J]. *J Web Semant*, **71**: 100662.
- Enkhsaikhan M, Holden E J, Duuring P, Liu W. 2021. Understanding ore-forming conditions using machine reading of text[J]. *Ore Geol Rev*, **135**: 104200.
- Falanga M, De Lauro E, Petrosino S, Rincon-Yanez D, Senatore S. 2022. Semantically enhanced IoT-oriented seismic event detection: An application to Colima and Vesuvius volcanoes[J]. *IEEE Int Things J*, **9**(12): 9789–9803
- Fan R, Wang L, Yan J, Song W, Zhu Y, Chen X. 2019. Deep learning-based named entity recognition and knowledge graph construction for geological hazards[J]. *ISPRS Int J Geo-Inf*, **9**(1): 15.
- García L F, Abel M, Perrin M, Dos Santos Alvarenga R. 2020. The GeoCore ontology: A core ontology for general use in geology[J]. *Comput Geosci*, **135**: 104387.
- Ge X T, Yang Y, Peng L, Chen L J, Li W C, Zhang W Y, Chen J H. 2022a. Spatio-temporal knowledge graph based forest fire

- prediction with multi source heterogeneous data[J]. *Remote Sens*, **14**(14): 3496.
- Ge X T, Yang Y, Chen J H, Li W C, Huang Z S, Zhang W Y, Peng L. 2022b. Disaster prediction knowledge graph based on multi-source spatio-temporal information[J]. *Remote Sens*, **14**(5): 1214.
- Guarino N. 1998. *Formal Ontology in Information Systems: Proceedings of the 1st International Conference, June 6-8, 1998, Trento, Italy*[M]. Amsterdam, Netherlands: IOS Press:4–10.
- Guarino N, Welty C. 2002. Evaluating ontological decisions with OntoClean[J]. *Commun ACM*, **45**(2): 61–65.
- Guichet X, Dubos-Sallée N, Cacas-Stentz M C, Rahon D, Martinez V. 2019. Efficient access to relevant knowledge extracted from geoscience literature dedicated to petroleum basin exploration by using IBM Watson[C]//*Proceedings of 2019 AAPG Annual Convention and Exhibition*. San Antonio: AAPG:87.
- Guizzardi G, Benevides A B, Fonseca C M, Porello D, Almeida J P A, Sales T P. 2022. UFO: Unified foundational ontology[J]. *Appl Ontol*, **17**(1): 167–210.
- Gutiérrez C, Sequeda J F. 2021. Knowledge graphs[J]. *Commun ACM*, **64**(3): 96–104.
- Hao X J, Ji Z, Li X H, Yin L Z Y, Liu L, Sun M Y, Liu Q, Yang R J. 2021. Construction and application of a knowledge graph[J]. *Remote Sens*, **13**(13): 2511.
- Hoerer T, Kuenzer C. 2022. SyntEO: Synthetic dataset generation for earth observation and deep learning: Demonstrated for offshore wind farm detection[J]. *ISPRS J Photogramm Remote Sens*, **189**: 163–184.
- Hogan A, Blomqvist E, Cochez M, D'Amato C, De Melo G, Gutierrez C, Kirrane S, Gayo J E L, Navigli R, Neumaier S, Ngomo A C N, Polleres A, Rashid S M, Rula A, Schmelzeise L, Sequeda J, Staab S, Zimmerman A. 2021. Knowledge graphs[J]. *ACM Comput Surv*, **54**(4): 71.
- Hu X M, Xu Y W, Ma X, Zhu Y, Ma C, Li C, Lü H, Wang X, Zhou C, Wang, C. 2023. Knowledge system, ontology, and knowledge graph of the Deep-Time Digital Earth (DDE): Progress and perspective[J]. *J Earth Sci*, **34**(5): 1323–1327.
- Jackson I. 2008. OneGeology: From concept to reality[J]. *Epis J Int Geosci*, **31**(3): 344–345.
- Kocoń J, Cichecki I, Kaszyca O, Kochanek M, Szydło D, Baran J, Bielaniec J, Gruza M, Janz A, Kanclerz K, Kocoń A, Kopyra B, Mielewszczenko-Kowszewicz W, Miłkowski P, Oleksy M, Piasecki M, Radliński L, Wojtasik L, Woźniak S, Kazienko P. 2023. ChatGPT: Jack of all trades, master of none[J]. *Inform Fusion*, **99**: 101861.
- Krisnadi A, Hu Y J, Janowicz K, Hitzler P, Arko R, Carbotte S, Chandler C, Cheatham M, Fils D, Finin T, Finin T, Ji P, Jones M, Karima N, Lehnert K, Mickel A, Narock T, O'Brien M, Raymond L, Shepherd A, Schildhauer M, Wiebe P. 2015. The GeoLink modular oceanography ontology[C]//*The Semantic Web-ISWC 2015: Proceedings of the 14th International Semantic Web Conference, Bethlehem, PA, USA, October 11–15, 2015, Proceedings, Part II* 14. Bethlehem: Springer International Publishing: 301–309.
- Laxton J L. 2017. Geological map fusion: OneGeology-Europe and INSPIRE[J]. *Spec Publ*, **408**(1): 147–160.
- Li Y S, Kong D Y, Zhang Y J, Tan Y H, Chen L. 2021. Robust deep alignment network with remote sensing knowledge graph for zero-shot and generalized zero-shot remote sensing image scene classification[J]. *ISPRS J Photogramm Remote Sens*, **179**: 145–158.
- Liu C, Chen J P, Li S, Qin T. 2022. Construction of conceptual prospecting model based on geological big data: A case study in Songtao-Huayuan area, Hunan Province[J]. *Minerals*, **12**(6): 669.
- Liu X H, Liu Z, Liu Y W, Tian J K. 2021. Integration of a geo-ontology-based knowledge model and spatial analysis into emergency response for geologic hazards[J]. *Nat Hazards*, **108**(2): 1489–1514.
- Liu X J, Zhang Y J, Zou H M, Wang F, Cheng X, Wu W P, Liu X Y, Li Y S. 2023. Multi-source knowledge graph reasoning for ocean oil spill detection from satellite SAR images[J]. *Int J Appl Earth Obs Geoinform*, **116**: 103153.
- Lv X, Xie Z, Xu D X, Jin X G, Ma K, Tao L F, Qiu Q, J Pan Y S. 2022. Chinese named entity recognition in the geoscience domain based on bert[J]. *Earth Space Sci*, **9**(3): e2021EA002166.
- Ma X G. 2022. Knowledge graph construction and application in geosciences: A review[J]. *Comput Geosci*, **161**: 105082.
- McBride B. 2004. The resource description framework (RDF) and its vocabulary description language RDFS[M]//*Handbook on Ontologies*. Berlin, Heidelberg: Springer: 51–65.
- Meyer L P, Stadler C, Frey J, Radtke N, Junghanns K, Meissner R, Dziwis G, Bulert K, Martin M. 2023. LLM-assisted

- knowledge graph engineering: Experiments with ChatGpt[C]//*First Working Conference on Artificial Intelligence Development for A Resilient and Sustainable Tomorrow*. Wiesbaden: Springer Vieweg: 103–115.
- Mezni H, Driss M, Boulila W, Atitallah S B, Sellami M, Alharbi N. 2022. Smartwater: A service-oriented and sensor cloud-based framework for smart monitoring of water environments[J]. *Remote Sens*, **14**(4): 922.
- Mughal M H, Shaikh Z A, Wagan A I, Khand Z H, Hassan S. 2021. ORFFM: An ontology-based semantic model of river flow and flood mitigation[J]. *IEEE Access*, **9**: 44003–44031.
- Oberle D. 2006. *Semantic Management of Middleware*[M]. New York: Springer: 33–53.
- Otte J N, Beverley J, Ruttenberg A. 2022. BFO: Basic formal ontology[J]. *Appl Ontol*, **17**(1): 17–43.
- Peters S E, Zhang C, Livny M, Ré C. 2014. A machine reading system for assembling synthetic paleontological databases[J]. *PLoS One*, **9**(12): e113523.
- Qu Y W, Perrin M, Torabi A, Abel M, Giese M. 2024. GeoFault: A well-founded fault ontology for interoperability in geological modeling[J]. *Comput Geosci*, **182**: 105478.
- Raskin R G, Pan M J. 2005. Knowledge representation in the semantic web for Earth and environmental terminology (SWEET)[J]. *Comput Geosci*, **31**(9): 1119–1125.
- Richard S M, Pearthree G, Aufdenkampe A K, Cutcher-Gershenfeld J, Daniels M, Gomez B, Kinkade D, Percivall G. 2014. Community-developed geoscience cyberinfrastructure[J]. *Eos*, **95**(20): 165–166.
- Rising J, Tedesco M, Piontek F, Stainforth D A. 2022. The missing risks of climate change[J]. *Nature*, **610**(7933): 643–651.
- Roldán-Molina G R, Ruano-Ordás D, Basto-Fernandes V, Méndez J R. 2021. An ontology knowledge inspection methodology for quality assessment and continuous improvement[J]. *Data Knowl Eng*, **133**: 101889.
- Santos N O, Rodrigues F H, Schmidt D, Romeu R K, Nascimento G, Abel M. 2024. O3PO: A domain ontology for offshore petroleum production plants[J]. *Expert Syst Appl*, **238**: 122104.
- Scherp A, Saathoff C, Franz T, Staab S. 2011. Designing core ontologies[J]. *Appl Ontol*, **6**(3): 177–221.
- Studer R, Benjamins V R, Fensel D. 1998. Knowledge engineering: Principles and methods[J]. *Data Knowl Eng*, **25**(1/2): 161–197.
- Sun S T, Dustdar S, Ranjan R, Morgan G, Dong Y S, Wang L Z. 2022. Remote sensing image interpretation with semantic graph-based methods: A survey[J]. *IEEE Sel Top Appl Earth Obs Remote Sens*, **15**: 4544–4558.
- Sure Y, Staab S, Studer R. 2009. Ontology engineering methodology[M]//*Handbook on Ontologies*. Berlin, Heidelberg: Springer: 135–152.
- Tang X M, Feng Z Q, Xiao Y T, Wang M, Ye T R, Zhou Y J, Meng J, Zhang B S, Zhang D W. 2023. Construction and application of an ontology-based domain-specific knowledge graph for petroleum exploration and development[J]. *Geosci Front*, **14**(5): 101426.
- Tiwari S, Al-Aswadi F N, Gaurav D. 2021. Recent trends in knowledge graphs: Theory and practice[J]. *Soft Comput*, **25**(13): 8337–8355.
- Wang B, Wu L, Xie Z, Qiu Q J, Zhou Y, Ma K, Tao L F. 2022. Understanding geological reports based on knowledge graphs using a deep learning approach[J]. *Comput Geosci*, **168**: 105229.
- Wang C B, Ma X G, Chen J G, Chen J W. 2018. Information extraction and knowledge graph construction from geoscience literature[J]. *Comput Geosci*, **112**: 112–120.
- Wang C S, Hazen R M, Cheng Q M, Stephenson M H, Zhou C H, Fox P, Shen S Z, Oberhänsli R, Hou Z Q, Ma X G, Feng Z Q, Fan J X, Ma, C, Hu X M, Luo, B. 2021a. The Deep-Time Digital Earth Program: Data-driven discovery in geosciences[J]. *Natl Sci Rev*, **8**(9): nwab027.
- Wang M H, Zhang X Y, Feng D E, Wang Y P, Tang W, Ye P. 2021b. Risk assessment of alpine skiing events based on knowledge graph: A focus on meteorological conditions[J]. *ISPRS Int J Geo-Inf*, **10**(12): 835.
- Wang S, Zhang X Y, Ye P, Du M, Lu Y X, Xue H N. 2019. Geographic Knowledge Graph (GeoKG): A formalized geographic knowledge representation[J]. *ISPRS Int J Geo-Inf*, **8**(4): 184.
- Wu J T, Orlandi F, O’Sullivan D, Dev S. 2022a. LinkClimate: An interoperable knowledge graph platform for climate data[J]. *Comput Geosci*, **169**: 105215.

- Wu J T, Orlandi F, O'Sullivan D, Pisoni E, Dev S. 2022b. Boosting climate analysis with semantically uplifted knowledge graphs[J]. *IEEE J Sel Top Appl Earth Obs Remote Sens*, **15**: 4708–4718.
- Yu T L, Yang H, Luo X W, Jiang Y F, Wu X, Gao J Q. 2021. Scientometric analysis of disaster risk perception: 2000–2020[J]. *Int J Env Res Pub He*, **18**(24): 13003.
- Zhang C, Govindaraju V, Borchardt J, Foltz T, Ré C, Peters S. 2013. GeoDeepDive: Statistical inference using familiar data-processing languages[C]//*Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data*. New York: Association for Computing Machinery: 993–996.
- Zhang M, Wang J L. 2022. Global flood disaster research graph analysis based on literature mining[J]. *Appl Sci*, **12**(6): 3066.
- Zhang X Y, Huang Y, Zhang C J, Ye P. 2022. Geoscience Knowledge Graph (GeoKG): Development, construction and challenges[J]. *Trans GIS*, **26**(6): 2480–2494.
- Zhao M W, Jiang L, Wang C, Yang C C, Yang X. 2021. On the topographic entity-oriented digital elevation model construction method for urban area land surface[J]. *Front Earth Sci Proc*, **15**(3): 580–594.
- Zheng K, Xie M H, Zhang J B, Xie J, Xia S H. 2022. A knowledge representation model based on the geographic spatiotemporal process[J]. *Int J Geogr Inf Sci*, **36**(4): 674–691.
- Zhou C H, Wang H, Wang C S, Hou Z Q, Zheng Z M, Shen S Z, Cheng Q M, Feng Z Q, Wang X B, Lv H R, Fan J X, Hu X M, Hou M C, Zhu Y Q. 2021. Geoscience knowledge graph in the big data era[J]. *Science China Earth Science*, **64**(7): 1105–1114.
- Zhou L, Cheatham M, Krisnadhi A, Hitzler P. 2020. Geolink data set: A complex alignment benchmark from real-world ontology[J]. *Data Intell*, **2**(3): 353–378.
- Zhu Y, Zhou W, Xu Y, Liu J, Tan Y. 2017. Intelligent learning for knowledge graph towards geological data[J]. *Sci Progr*, **2017**: 5072427.