SCIENCE CHINA
Earth Sciences

Geoscience knowledge graph in the big data era

Chenghu ZHOU, Hua WANG, Chengshan WANG[3], Zengqian HOU[4], Zhiming ZHENG[5], Shuzhong SHEN[6], Qiuming CHENG[3], Zhiqiang FENG[7], Xinbing WANG[8], Hairong LV[9], Junxuan FAN[6], Xiumian HU[6], Mingcai HOU[10] and Yunqiang ZHU[10]

Citation: SCIENCE CHINA Earth Sciences **64**, 1105 (2021); doi: 10.1007/s11430-020-9750-4

View online: https://engine.scichina.com/doi/10.1007/s11430-020-9750-4

View Table of Contents: https://engine.scichina.com/publisher/scp/journal/SCES/64/7

Published by the Science China Press

SCIENCE CHINA
Earth Sciences

•RESEARCH PAPER•

July 2021 Vol.64 No.7: 1105–1114
https://doi.org/10.1007/s11430-020-9750-4

Geoscience knowledge graph in the big data era

Chenghu ZHOU[1,2*], Hua WANG[1,2†], Chengshan WANG[3], Zengqian HOU[4], Zhiming ZHENG[5], Shuzhong SHEN[6], Qiuming CHENG[3], Zhiqiang FENG[7], Xinbing WANG[8], Hairong LV[9], Junxuan FAN[6], Xiumian HU[6], Mingcai HOU[10] & Yunqiang ZHU[1,2]

[1] *State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China;*
[2] *University of Chinese Academy of Sciences, Beijing 100049, China;*
[3] *School of Earth Sciences and Resources, China University of Geosciences (Beijing), Beijing 100083, China;*
[4] *Key Laboratory of Deep-Earth Dynamics of Ministry of Natural Resources, Institute of Geology, Chinese Academy of Geological Sciences, Beijing 100037, China;*
[5] *Institute of Artificial Intelligence, Beihang University, Beijing 100191, China;*
[6] *State Key Laboratory for Mineral Deposits Research, School of Earth Sciences and Engineering, Nanjing University, Nanjing 210023, China;*
[7] *Sinopec Petroleum Exploration & Production Research Institute, Beijing 100083, China;*
[8] *School of Electronic, Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China;*
[9] *Department of Automation, Tsinghua University, Beijing 100084, China;*
[10] *Institute of Sedimentary Geology, Chengdu University of Technology, Chengdu 610059, China*

Received November 30, 2020; revised January 4, 2021; accepted March 1, 2021; published online May 25, 2021

Abstract Since the beginning of the 21st century, the geoscience research has been entering a significant transitional period with the establishment of a new knowledge system as the core and with the drive of big data as the means. It is a revolutionary leap in the research of geoscience knowledge discovery from the traditional encyclopedic discipline knowledge system to the computer-understandable and operable knowledge graph. Based on adopting the graph pattern of general knowledge representation, the geoscience knowledge graph expands the unique spatiotemporal features to the Geoscience knowledge, and integrates geoscience knowledge elements, such as map, text, and number, to establish an all-domain geoscience knowledge representation model. A federated, crowd intelligence-based collaborative method of constructing the geoscience knowledge graph is developed here, which realizes the construction of high-quality professional knowledge graph in collaboration with global geo-scientists. We also develop a method for constructing a dynamic knowledge graph of multi-modal geoscience data based on in-depth text analysis, which extracts geoscience knowledge from massive geoscience literature to construct the latest and most complete dynamic geoscience knowledge graph. A comprehensive and systematic geoscience knowledge graph can not only deepen the existing geoscience big data analysis, but also advance the construction of the high-precision geological time scale driven by big data, the compilation of intelligent maps driven by rules and data, and the geoscience knowledge evolution and reasoning analysis, among others. It will further expand the new directions of geoscience research driven by both data and knowledge, break new ground where geoscience, information science, and data science converge, realize the original innovation of the geoscience research and achieve major theoretical breakthroughs in the spatiotemporal big data research.

Keywords Geoscience knowledge graph, All-domain geoscience knowledge representation model, Federated crowd intelligence collaboration, High-precision geological time scale

* Corresponding author (email: zhouch@lreis.ac.cn)
† Corresponding author (email: wangh@lreis.ac.cn)

Citation: Zhou C, Wang H, Wang C, Hou Z, Zheng Z, Shen S, Cheng Q, Feng Z, Wang X, Lv H, Fan J, Hu X, Hou M, Zhu Y. 2021. Geoscience knowledge graph in the big data era. Science China Earth Sciences, 64(7): 1105–1114, https://doi.org/10.1007/s11430-020-9750-4

1. Introduction

A new round of scientific and technological revolution and industrial transformation is emerging around the world nowadays. Human beings are stepping into a new era of science and technology featuring the integrated development of big data, cloud computing, artificial intelligence, blockchain, and Internet of Things. The second-generation artificial intelligence based on big data and deep learning has been widely and successfully applied in image recognition, language translation, and other fields (Guo et al., 2014; Guo, 2017a, 2017b). Big data is not only changing the way of human life, production, and thinking, but also guiding the scientific research from the scientific paradigm of experiment, theory, and simulation to the fourth scientific research paradigm driven by big data (Tansley and Tolle, 2009).

The modern earth observation system realizes hour-level global monitoring, and various automatic observation station networks can continuously observe global precipitation, biomass, and other elements of the earth system, and digital publication can be accessed anytime and anywhere. These changes indicate the geoscience research has entered an era of big data with global coverage, all-weather monitoring, and all-element observation. At the same time, as a typical data-intensive science, geoscience faces challenges in data integration and sharing, data mining and knowledge discovery, such as data chaos and lack of mechanisms in the spatial statistical analysis. In addition, many potential advantages of big data have not been brought into full play in geoscience-related researches, and there is an urgent need to develop theories and methods of knowledge-driven big data analysis in geoscience. Therefore, to construct the all-domain geoscience knowledge graph (GKG) and explore the geoscience knowledge evolution are the frontier areas and strategic focus of the contemporary researches on geoscience knowledge discovery (Zhai et al., 2018; Wang et al., 2019). Artificial intelligence is the key to the value mining and promotion of big data, while knowledge graph is one of the important cornerstones of artificial intelligence and the core foundation of integrating statistical representation and physical representation. This paper expounds and discusses the key scientific issues and frontier directions of geoscience knowledge graph research oriented towards geoscience big data (hereinafter referred to as geo-big data) analysis from the aspects of graph pattern of geoscience knowledge representation, the method of constructing the geoscience knowledge graph and the application of geoscience knowledge graph, and looks into the future researches of spatio-temporal big data analysis and knowledge discovery in geoscience.

2. Geoscience knowledge representation model

Geoscience is a science that studies the formation, evolution, and interaction of various spheres of Earth, including the atmospheric science, marine science, geography, geology, geophysics, and others, which has complex and diverse disciplinary knowledge systems (Sun, 2017). Therefore, the multi-scale spatiotemporal features of geoscience phenomena and processes, integrating various information carriers such as map, text and number, and the geoscience knowledge representation models spanning all discipline branches of geoscience constitute the basis and starting point of researches on geoscience knowledge graph.

2.1 Geoscience knowledge system and knowledge graph

As a complex giant system, the earth system stretches across millions of miles in space and billions of years in time. It is a systematic knowledge project and the unremitting pursuit of geo-scientists to build a complete system of geoscience knowledge. Geoscience knowledge follows the general characteristics of the common knowledge. For examples, the Organization for Economic Cooperation and Development (OECD) classifies knowledge into four categories, namely, knowledge about things and reality, knowledge about natural laws and principles, knowledge about skills and know-how, and knowledge about human resources (OECD, 1996); Benjamin Bloom, the US psychologist and educator, classifies knowledge into factual knowledge, conceptual knowledge, procedural knowledge, and meta knowledge (OECD, 1996). All these existing research achievements related to knowledge engineering can be the foundation of constructing the geoscience knowledge system.

However, compared with other disciplines, multi-scale spatiotemporal features are the basic elements of geoscience knowledge. On one hand, when geo-scientists think about and conduct researches, they usually focus on the specific spatial scope and time span of the research objects. For example, geologists often use one million years as the basic time unit to measure major geological events, and use the intercontinental spatial scale in studies of plates motion. On the other hand, due to technical limitation, the observation and analytical methods used in geoscience are also various

ways. For example, paleontologists' fossil specimens and geo-chronologists' gold spike section have limited and observable spatial scales. Therefore, the core and innovation of the geoscience knowledge system construction is to develop an open and expandable geoscience knowledge framework by combining the general knowledge classification system and the spatiotemporal features of various element carriers of map, text and number. The framework is developed by using tree-like knowledge architecture as the main body supplemented by the network and knowledge chain architecture.

Geoscience knowledge graph is a kind of computer-understandable and calculable knowledge system in which relevant knowledge is effectively organized through a structured graph pattern. The concept and prototype of knowledge graph can be traced back to the 1960s, and it has been widely applied in the field of library and information science. As one of pioneering work, Boyack et al. (2005) constructed a knowledge graph with nodes and edges to reveal the internal relationships and disciplinary relationships among 800,000 scientific articles. Auer et al. (2007) promoted the construction of an open linked database, DBpedia, by adopting semantic network and other methods. The method of describing the concepts, entities and their relationships based on semantic network is also an effective representation of the geoscience knowledge graph (Tang, 2020; Zhang et al., 2020). In 2012, Google officially launched the Knowledge Graph Engine (Singhal, 2012), a new-generation knowledge graph "Knowledge Vault" for obtaining factual information from unstructured network texts (Dong et al., 2014; Lu et al., 2017). The Knowledge Vault containing more than 600 million entities and more than 18 billion attribute or relationship nodes greatly promoted the development and application of knowledge graph technology and methods.

Since 2017, Open Knowledge Network had been brought into US national science and technology strategy (NSTC, 2018). In 2019, the US National Science Foundation (NSF) funded 43 pilot projects with a total budget of $39 million for accelerating the integration of disciplines, 21 projects of them are related to Open Knowledge Network, and more investment will be put on this theme in the future (https://www.nsf.gov/pubs/2019/nsf19050/nsf19050.jsp; https://www.nsf.gov/od/oia/convergence-accelerator/index.jsp). The Deep Time Knowledge Graph project leaded by the University of Idaho introduced the latest semantic web model to build a machine-readable deep time "language" to connect the international and regional geochronology standards, to manage the concept of times in different versions of the geochronological scale, and use it to explore and analyze the deep time data under the network environment (Ma et al., 2020). At present, SocialWiki, together with Douban Time and other youth media, jointly launched the "Human Knowledge Pedigree Construction Project", attempting to develop a broader graph of disciplines through global collaboration to help people explore and learn knowledge. These studies and practices will contribute to promoting the research on geoscience knowledge graph.

2.2 Self-adaptive representation model of all-domain geoscience knowledge graph

Knowledge representation is the basis for constructing the computer-understandable and calculable knowledge graph, and an important step in the process of knowledge communication. The common ways of knowledge representation include natural languages, structured tables, graphics and images, etc. Natural language representation is generally qualitative and vague description, structured tables cannot sufficiently describe the spatiotemporal relationships between geoscience entities (hereinafter refer to as geo-entities), and graphics and images are not good at accurately describing complex geoscience processes. Therefore, there is an urgent need to develop a comprehensive formal representation language (Xu et al., 2010). The graph pattern represented by the directed graph effectively establishes the semantic association between knowledge entities, being a basic model for geoscience knowledge representation. This basic model needs expanding both in time and space to describe the geoscience knowledge. For example, in the field of geography, scholars have put forward YAGO2 (Hoffart et al., 2013), GeoKG (Wang et al., 2019) and other geographic knowledge representation models to record the geographic knowledge in a directed graph structure (Ballatore et al., 2015). YAGO2, upgraded version of the general knowledge graph YAGO, adds the predicative description of temporal and spatial expression to record the temporal and spatial information of each entity. GeoKG puts forward a geographic knowledge representation model to describe the evolution process of geographic entities. Time, space, attribute, state, change and relationship are also taken into consideration.

The existing geoscience knowledge representation models which simply supplement or record the temporal, spatial and part of the attribute information are difficult to represent the complex all-domain geosciences knowledge (Zhang et al., 2020; Oramas et al., 2017). It is a challenging task to construct the representation model of geoscience knowledge graph across spatiotemporal dimensions from the perspective of the essence of graph structure by integrating the complex spatiotemporal features, computational attributes and the relationships and rules of geoscience knowledge. Here we propose a basic model of the self-adaptive representation of all-domain geoscience knowledge graph as shown in Figure 1. This fundamental model is composed of complex spatiotemporal information representation model (entity object representation model) and geo-entity object relationship

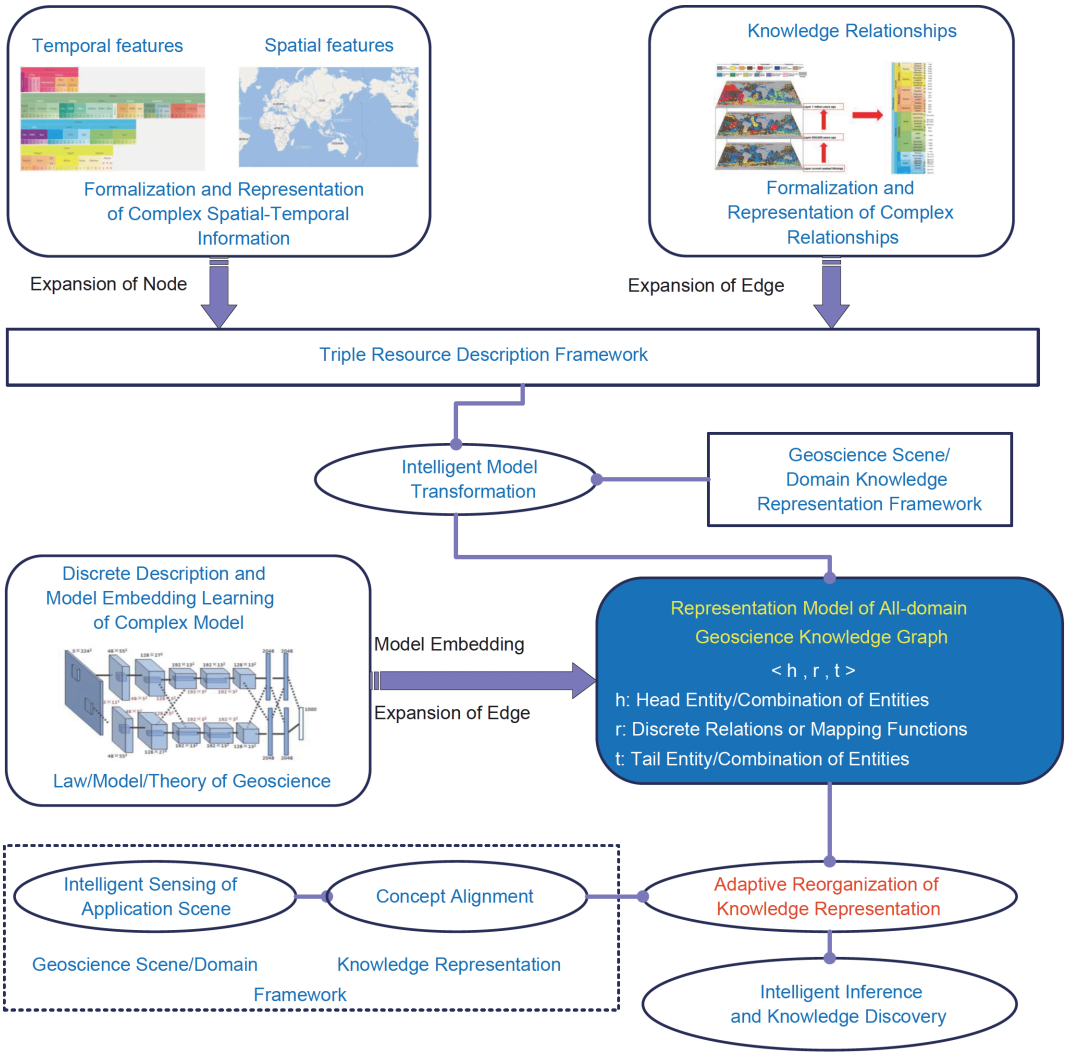

Figure 1 Self-adaptive representation model of all-domain geoscience knowledge graph.

(edge). The model still adopts the triples of Resource Description Framework (RDF) of head entity/entity combination, tail entity/entity combination and discrete relationship/mapping functions. And more, three aspects of expansion are taken to include spatiotemporal features on the nodes, complex model into the edge relationships, and the intelligent recognition and concept alignment to the geoscience scenes. This expanded model can achieve the self-adaptive representation of Geoscience knowledge graph in the whole domain.

3. Methods of constructing the geoscience knowledge graph

Systematically summarizing the existing scientific knowledge of human beings has always been one of the components of scientific research activities, such as compiling academic classics and dictionaries, encyclopedias, and so on. In the 1960s, Price, the founder of scientometrics, discovered the exponential growth law of scientific knowledge by using statistical methods based on the database of scientific papers, which became one of the pioneering works in the application of knowledge graph (Chen et al., 2008a).

3.1 Strategies and methods for constructing knowledge graph

Generally, strategies for constructing knowledge graph can be divided into "bottom-up" and "top-down" approaches. The construction of open knowledge graph mostly adopts the "bottom-up" approach, which automatically extracts concepts or entities from various text data as well as the relationships between them, such as with the Google's Knowledge Vault. Knowledge graphs in professional fields mostly adopt the "top-down" approach, which determines

the ontology and entities to be constructed in advance, and then constructs them in a professional way, so that the constructed knowledge graphs are of high professional level. In the "Deep Time Digital Earth" big science project (hereinafter referred to as DDE) initiated and being implemented by Chinese scientists, 18 working groups from different disciplines have been set up. DDE promotes the construction of disciplinary knowledge systems and knowledge rules of solid earth science by adopting the "top-down" approach, with the purpose of providing a foundation for the next step in the construction of the global geoscience knowledge graph.

As an emerging research field, methods for constructing the knowledge graph are developed rapidly, and the automatic and intelligent construction methods constantly emerge, which greatly enhances the ability of knowledge graph construction. The conventional methods based on manual input and editing are still widely adopted and constantly improved. In particular, the level of automation and intelligence of expert knowledge alignment is improving. For example, Cyc system (Lenat, 1995), Freebase and Wikidata are based on crowd sourcing and web collaborative editing (Chah, 2017; Mitraka et al., 2015), while DBpdeia, YAGO, NELL, and PROSPERA extract knowledge from large-scale, semi-structured or unstructured text data (Ansari et al., 2019; Carlson et al., 2010; Nakashole et al., 2011). These methods and systems provide a solid foundation for constructing geoscience knowledge graph.

Geoscience knowledge graph based on the well-known knowledge system has a clear and definite explanation of all knowledge nodes in the field of geoscience (including well-known phenomena and facts, basic concepts and definitions, natural principles and laws, observations and analytical methods, etc.) as well as the relationships between the knowledge nodes. It is a machine-understandable geoscience knowledge base and inference engine. Human-machine collaborative editing and importing strategies, such as crowd intelligence collaborative construction method, are mostly adopted for the relatively stable and mature knowledge systems, especially for factual knowledge and conceptual knowledge in the field. As to the knowledge scattered in a large volume of literature, especially those published academic papers, books and research reports, text literature data mining and knowledge discovery methods are often adopted, such as the network text analysis and knowledge discovery methods. With the further development of knowledge graph, domain expert knowledge and dynamic knowledge from web texts will be integrated to form a hybrid construction system.

3.2 Geoscience knowledge graph construction through crowd intelligence collaboration

The geoscience knowledge system covers a wide range, and most of the knowledge and experience come from the geoscience experts. Manually inputting the knowledge and experience of well-known experts with traditional methods requires a high degree of collaboration and great cost. It is extremely difficult to formally represent the geoscience knowledge with high uncertainty or ambiguity, therefore the progress of GKG construction in this area is rather slow. With the rapid development and penetration of intelligent mobile terminals, the crowd intelligence collaboration based on mobile internet opens a new way to solve the above problems. The collective forces and the collection of structured patterns can update the geoscience knowledge graph in real time, and verify the knowledge mined by machine with the support of crowd intelligence. Here we proposes a federated crowd intelligence knowledge graph construction framework with expert knowledge as the core (Figure 2).The following key technical issues need to be addressed.

(1) The contradictions and conflicts among knowledge from different experts will inevitably arise in the large-scale geoscience knowledge graph constructed through crowd intelligence collaboration. It needs to effectively identify the iterative updating of new and old knowledge and the collision of views between different theoretical systems. For examples, Gangemi et al. (2007) proposed the Collaborative Ontology-Design model to discover and handle conflicts of the process of the knowledge creation, integration and collaborative development in the heterogeneous semantic environment; Chen et al. (2008b) divided the conflicts in the knowledge collaborative construction into three categories of collaborative conflicts, or consistent conflicts and abnormal conflicts, and developed the Command Package Pool method to realize the automatic detection of conflicts. In the framework of Figure 2, the conflict knowledge fusion and correction method based on the knowledge contribution and credibility evaluation, are adopted to realize the automatic conflict detection of geoscience knowledge created by crowd intelligence.

(2) The research and establishment of a sustainable co-construction and sharing model is an important guarantee mechanism for the construction of the global geoscience knowledge graph. If knowledge owners unwilling disclose the details of knowledge when they participate in the construction of knowledge graph through crowd intelligence, it is necessary to develop knowledge storage proof methods with protection ability of privacy and intellectual property rights. And a knowledge-sharing model called as "I am the master of my knowledge" is preferred. In order to motivate scientists to actively participate in the process of construction of knowledge graph through crowd intelligence collaboration, it is also necessary to establish the evaluation method of knowledge value contribution and the credibility evaluation model of knowledge contributors under the crowd intelligence collaboration, so as to realize the evaluation of

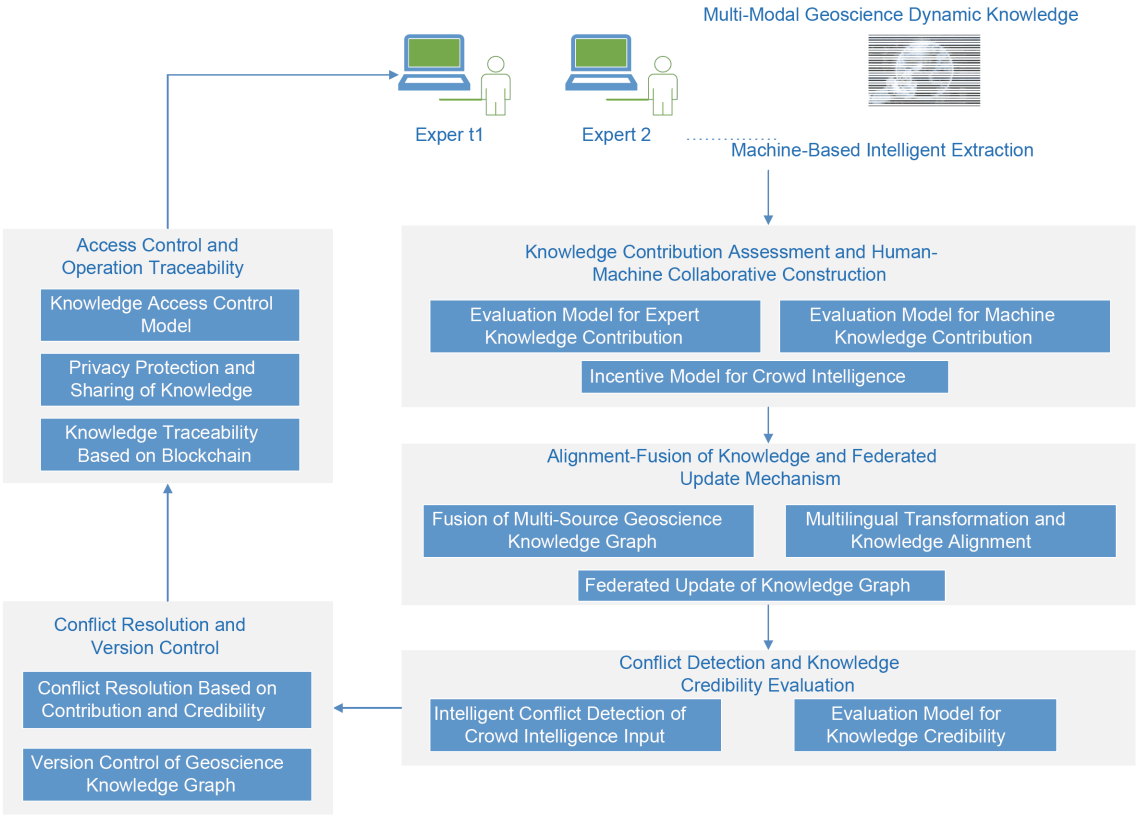

Figure 2 Construction framework of geoscience knowledge graph through federated crowd intelligence collaboration.

expert knowledge contribution and establish an effective incentive mechanism of crowd intelligence collaboration.

(3) The large-scale collaboration of global geo-scientists needs to effectively solve the problem of multilingual representation of knowledge, and to study and establish consistency-based fusion technologies and methods of multi-source knowledge graph based on rules, statistics and deep learning.

3.3 Construction of dynamic knowledge graph of multi-modal geoscience data based on in-depth analysis

Massive structured and unstructured geoscience literature (hereinafter referred to as geo-literature) which are published or only internally used contain a huge amount of geoscience knowledge, especially the latest dynamic knowledge. It is important to construct the dynamic geoscience knowledge graph by using machine learning, deep learning, and other modern information technology to extract and update geoscience knowledge from these massive text literature once data perception, entity recognition, and relationship extraction are solved, respectively. Here we propose a method to construct dynamic knowledge graph of multi-modal geoscience data based on in-depth analysis (Figure 3): Firstly, unstructured data perception is conducted by using in-depth analysis of multi-source geoscience data. To realize text association and multi-source data perception, massive unstructured text materials such as text, pictures, data tables, and maps are classified, and the associated attributes from the same source such as map name, region, cartographic index, latitude and longitude range, time, are labeled. Based on these labels, text segmentation, plain text extraction and syntactic analysis will be carried out. Especially, the non-substantive semi-structured text is eliminated by adopting some existing rule knowledge. With all these processing, graphs, texts and numbers from different sources with certain similarities are labeled and associated by adopting the text matching and statistical learning method, especially the keyword information in the text description need be extracted through rule-based filtering and neural network model. Compared with the general text source, the geoscience literature often contains a large number of maps, tables, and other professional components. Consequently, in the process of knowledge acquisition, we need to conduct in-depth analysis of maps, understand the meaning of the map symbols, identify various kinds of spatial relationships, and then sort out geo-objects in specific texts combining with the geoscience knowledge system.

Secondly, entity object and knowledge are extracted by using keywords. One of the keys in deep learning is the high-

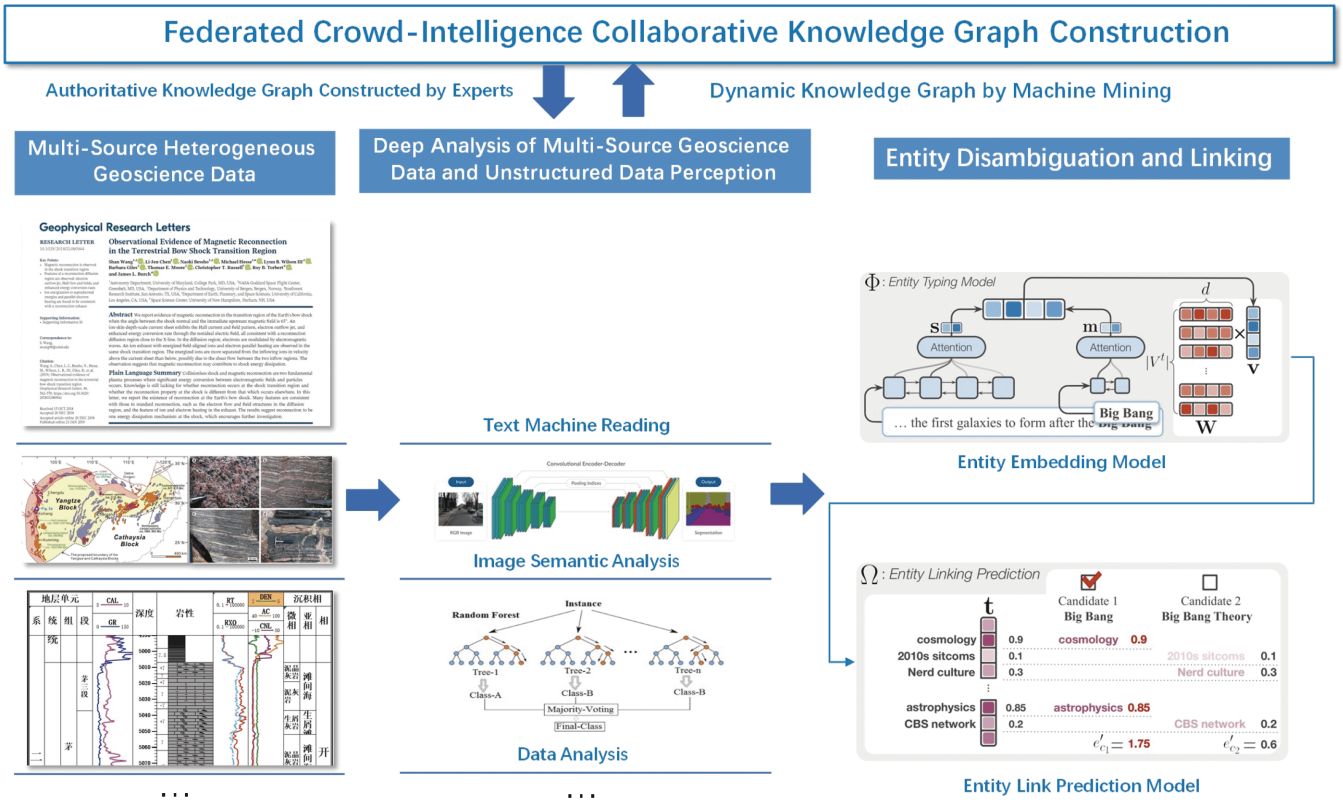

Figure 3 Construction framework of dynamic knowledge graphs of multi-modal geoscience data based on in-depth analysis.

quality training corpus which is difficult to be realized by manual selection and annotation. Therefore, it is crucial to develop efficient and credible unsupervised learning algorithm, such as object extraction based on keyword graph model. The algorithm first quantifies the statistical characteristics of geo-entities based on the word segmentation results of massive texts and the Term Frequency and Inverse Document Frequency (TF-IDF) algorithm. Then it builds the list of common words according to the sorted results and constructs the language network graph of massive texts. Based on the graph embedding vectors and the similarity among the word embeddings, it searches for important words or phrases in the language network graphs and screens out keywords in the texts as well as extracts the corresponding object entities and constructs the index relationships among graphs, text description, numbers and keywords. Finally, the matching among the above elements in the geoscience knowledge is completed to realize the geoscience knowledge extraction based on unstructured texts.

Thirdly, the disambiguation of knowledge and the construction of dynamic geoscience knowledge graphs are conducted. To solve the ambiguity and conflicts of geoscience knowledge caused by multiple data sources, classification and clustering are conducted through deep reinforcement learning using the specific spatiotemporal semantic association of entity concepts as features. In this way, we could address the problems of polysemy and synonymy. Further, it completes attribute alignment to eliminate knowledge conflicts by training the information source credibility through feature learning, and the source attributes as features.

4. Typical applications of geoscience knowledge graph

The wide applications of geoscience knowledge graph can promote the integration of geoscience, information science, and data science, and can accelerate the development of these disciplines. Knowledge-driven spatiotemporal analysis of big data is helpful to implement more accurate analysis of geo-big data, and promote the comprehensive analysis by combining the statistical and physical characterizations of geo-big data. Based on the existing geoscience knowledge base and knowledge engine, it can promote the research of geoscience knowledge system, and understand the evolution characteristics of geoscience knowledge and even discover new geo-knowledge with breakthroughs and innovations. The integration of geoscience knowledge and cartography knowledge can promote the intelligent and automatic de-

velopment of map making; the combination of geoscience knowledge with earth system models may push the exploration and prediction research of mineral resources. The following paragraphs focus on these three typical application areas of geoscience knowledge graph.

4.1 Geoscience knowledge reasoning

Geoscience knowledge reasoning refers to the establishment of new relationships among geo-entities, the understanding of the characteristics of evolution of geoscience knowledge system, as well as the discovery of new geoscience knowledge through computer reasoning from the relationships between entity concepts in geoscience knowledge graph. At present, the commonly used methods for knowledge reasoning include symbolic reasoning and statistical reasoning. The core of symbolic reasoning is to use rule of relevance to infer new entity relationships from existing ones and detect possible logical conflicts within them, such as the text-enhanced knowledge embedding method named TEKE proposed by Wang and Li (2016). Machine learning and other methods are adopted in statistical reasoning to develop new relationships among entities in knowledge graph based on past experience and analysis and verify or infer assumptions using maximum-a-posteriori (MAP) and other statistical methods, like Parallel Universe TransE (puTransE), an adaptable and robust transition matrix model proposed by Tay et al. (2017) for learning entity relationships on knowledge graph.

Neither symbolic reasoning nor statistical reasoning can effectively convey the spatiotemporal dependence and non-stationarity (heterogeneity) of geoscience knowledge, and they fail to make full use of the relationships between multi-form features of big data in geosciences to set up model. So it is difficult to interpret and predict multiple attributes of geo-knowledge. A conceptual model of bidirectional reasoning is proposed based on entity attributes and spatiotemporal links to solve this problem. In view of the fact that the evolution of geo-entities occurs over a long-time span and involves many time node information, causal convolution and expansion factor are used to achieve the learning of features of historical labeled information over a long period. It can avoid the vanishing gradient or gradient explosion problem that the traditional recurrent neural networks (RNNs) may incur when dealing with long sequence input. According to the spatial relationships of geo-entities and their corresponding topological models, a learning mechanism of spatial characteristics of geo-entities can be established based on heterogeneous graph construction and graph convolution neural network. Using the dynamic spatiotemporal relationships of the multi-task learning model, it can extract the deep temporal and spatial features of geo-entities and achieve the bidirectional correlations between attributes and spatiotemporal correlations and deduce the spatiotemporal linkage from attributes.

4.2 Construction of high-resolution geological time scale based on geoscience knowledge graph

An unified spatiotemporal framework is the prerequisite and cornerstone for geoscience research, being a geoscientist's dream to establish an unified deep-time framework with ten-thousand-year time resolutions. Golden spike defined with the first appearance of fossils is often called in question (Aubry et al., 1999; Walsh et al., 2004; Lucas, 2018; Davydov, 2020). An attempt has been made to analyze the ocean of data from stratigraphy, paleontology, geochronology, astronomical cycle and isotope chronology and other related disciplines with big data analysis techniques. This research also needs the support of a specific geoscience knowledge graph, which will launch the post-stratotype stage of international stratigraphy research based on traditional "Golden spike/stratotype" and geological chronology. Two main tasks must be conducted to achieve the above aim.

Firstly, intelligent methods for automatically extracting basic concepts, terms, specifications, technical methods and their interrelations of stratigraphy, paleontology, geochronology, astronomical cycle and isotope chronology is to be developed. Coding specifications suitable for describing geoscience knowledge graphs based on international standards and specifications, and machine-understandable ontology of knowledge graphs of stratigraphy, paleontology, geochronology, astronomical cycle and isotope chronology, especially time ontology should be established.

Secondly, artificial intelligence algorithms based on high performance computing and artificial intelligence technology are developed to calibrate time and time rate, determine major biological environmental events. It is necessary to explain geo-knowledge nodes (including basic concepts, objects, phenomena, processes, standards, methods, etc.) and their mutual relationships in a clear and definite manner and establish the geological time scale of different geological periods step by step, such as the new generation of geological time scale of 100,000 to 10,000 years in Paleozoic. Finally, a geological time scale spanning the entire evolution history of Earth will be formed.

4.3 Intelligent map editing and mapping

As a visual representation and transmission of spatial information, map and Cartography has a long history and has evolved with the progress in understanding the human cognition and the development of science and technology (Guo and Ying, 2017; Chen and Chen, 2018). Map-making is a complex, time-consuming and labor-intensive task, involving many steps such as map design, map editing, and map

generalization. The research and development of automatic and intelligent mapping system has always been the focus in the field of cartography. With the development of symbolic artificial intelligence (the first generation of artificial intelligence), there was a surge of studying the mapping expert systems from the early 1980s to the mid-1990s (Xiong, 2019). However, the bottleneck for knowledge engineering of cartography remains unsolved. Geoscience knowledge graph may light the smart map-making.

The intelligent mapping based on geoscience knowledge graphs is driven by geo-knowledge, cartographic knowledge and intelligent selection of big data under the constraint of the basic laws and rules of cartography. It implements autonomous judgment and collocation of various mapping resources including mapping data, models, templates and methods, and the map product quality and optimization strategy. The core tasks of intelligent map-making include the external task of mapping workflow determination by knowledge-driven and the internal task of map making assisted by data intelligence. The external task driven by the workflow knowledge will determine the whole mapping process, while the internal task includes a series of activities such as data processing, overall design, content organization, symbolic representation, map output, etc. The internal task is implemented with the guidance of the external task step by step.

5. Conclusions

Geoscience knowledge graph is one of the frontiers of geoscience researches. Building geoscience knowledge graph is a systematic knowledge project for the community of geosciences. An adaptable representation model of geoscience knowledge graph which was proposed in this paper is composed of complex spatiotemporal information representation model (entity object representation model) and geo-entity object relationship (edge). This model provides the basic architecture for building all-domain geoscience knowledge graph. Federated crowd intelligence collaboration method and dynamic abstraction method from multi-form geo-big data based on deep analysis are two different strategies and key techniques for establishing geoscience knowledge graph. Moreover, this paper illustrates three typical applications of geoscience knowledge graph from geo-knowledge reasoning, high-resolution geological time scale construction to intelligent mapping. These cases provide the idea and guidance for further applications of geoscience knowledge graph.

As a new geoscience research direction, knowledge graph is just in the early stages and need more in-depth studies such as structural organization and semantic representation of geoscience knowledge architecture, the large-scale knowledge graph construction and the big data analysis and application. In particular, the following core issues need to be studied urgently.

(1) The representation of geoscience knowledge integrated with map elements. At present, knowledge representation mostly adopts the basic architecture model of "node-edge", which cannot effectively integrate various thematic maps of geoscience research, and fails to analyze the spatial-temporal relationships of complex geo-phenomena or events effectively. The extended model proposed in this paper still face the challenges in implementing intelligent extraction and analysis of map elements.

(2) Rapid reconstruction of knowledge graph for different application theme. Nowadays, most knowledge graphs are constructed in existing scenes. The tough challenges in the application of knowledge graphs are how to construct knowledge graphs with "time-space-theme" as the core and achieve efficient reconstruction and adaptive analysis.

(3) Knowledge discovery of multivariate big data and updating of geoscience knowledge graph. Currently, the dynamic knowledge graph is constructed mostly based on literature, and it is difficult to discover knowledge from more sources and check the reliability and credibility of new knowledge as well as update the knowledge graph quickly.

Acknowledgements *This paper benefitted from the guidances and supports from many domestic and overseas colleagues. The theme of this paper comes from our collective thinking. So this paper should be a piece of collective work. The authors want to specially thank the experts and scholars who participated in the Shuang Qing Forum themed "Data-driven Geoscience: from Tradition to the Data Age" in 2019 and Deep Time Digital Earth Series Symposiums. The authors appreciate Prof. Yupeng Yao and Chaolin Zhao from the Geoscience Department of NSFC for their guidance on the overall design and research directions in the original exploration project of NSFC. Many thanks are also given to the anonymous peer reviewers for their valuable comments and suggestions. This work was supported by the National Natural Science Foundation of China (Grant Nos. 41421001, 42050101, and 42050105).*

References

Ansari G A, Saha A, Kumar V, Bhambhani M, Sankaranarayanan K, Chakrabarti S. 2019. Neural program induction for KBQA without gold programs or query annotations. In: Proceedings of the 28th International Joint Conference on Artificial Intelligence. Macao: AAAI Press. 4890–4896

Aubry M P, Berggren W A, Van Couvering J A, Steininger F. 1999. Problems in chronostratigraphy: Stages, series, unit and boundary stratotypes, global stratotype section and point and tarnished golden spikes. Earth-Sci Rev, 46: 99–148

Auer S, Bizer C, Kobilarov G, Lehmann J, Cyganiak R, Ives Z. 2007. DBpedia: A nucleus for a web of open data. In: Aberer K, Choi K-S, Noy N, Allemang D, Lee K-I, Nixon L, Golbeck J, Mika P, Maynard D, Mizoguchi R, Schreiber G, Cudré-Mauroux P, eds. The Semantic Web. ISWC 2007, ASWC 2007. Lecture Notes in Computer Science, vol 4825. Berlin, Heidelberg: Springer. 722–735

Ballatore A, Bertolotto M, Wilson D. 2015. A structural-lexical measure of semantic similarity for geo-knowledge graphs. ISPRS Int J Geo-Inform, 4: 471–492

Boyack K W, Klavans R, Börner K. 2005. Mapping the backbone of science. Scientometrics, 64: 351–374
Carlson A, Betteridge J, Kisiel B, Settles B, Hruschka Jr. E R, Mitchell T M. 2010. Toward an architecture for never-ending language learning. In: Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2010. Atlanta, Georgia, USA, July 11–15, 2010. 1306–1313
Chah N. 2017. Freebase-triples: A methodology for processing the freebase data dumps
Chen J, Chen J. 2018. GlobeLand30: Operational global land cover mapping and big-data analysis. Sci China Earth Sci, 61: 1533–1534
Chen Y, Liu Z, Chen J, Hou J. 2008a. History and theory of mapping knowledge domains (in Chinese). Stud Sci Sci, 26: 449–460
Chen Y, Zhang S, Peng X, Zhao W. 2008b. A collaborative ontology construction tool with conflicts detection, In: Fourth International Conference on Semantics, Knowledge and Grid. Los Alamitos: IEEE Computer Society. 12–19
Davydov V I. 2020. Shift in the paradigm for GSSP boundary definition. Gondwana Res, 86: 266–286
Dong X, Gabrilovich E, Heitz G, Horn W, Lao N, Murphy K, Strohmann T, Sun S, Zhang W. 2014. Knowledge vault: A web-scale approach to probabilistic knowledge fusion. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 601–610
Gangemi A, Presutti V, Catenacci C, Lehmann J, Nissim M. 2007. C-ODO: An OWL meta-model for collaborative ontology design. In: Proceedings of the Workshop on Social and Collaborative Construction of Structured Knowledge (CKC 2007) at the 16th International World Wide Web Conference (WWW2007). Banff
Guo H D. 2017a. Big data drives the development of Earth science. Big Earth Data, 1: 1–3
Guo H D. 2017b. Big Earth data: A new frontier in Earth and information sciences. Big Earth Data, 1: 4–20
Guo H D, Wang L, Chen F, Liang D. 2014. Scientific big data and Digital Earth. Chin Sci Bull, 59: 5066–5073
Guo R, Ying S. 2017. The rejuvennation of cartograph in ICT era (in Chinese). Acta Geodaet Cartograph Sin, 46: 1274–1283
Hoffart J, Suchanek F M, Berberich K, Weikum G. 2013. YAGO2: A spatially and temporally enhanced knowledge base from Wikipedia. Artificial Intell, 194: 28–61
Lenat D B. 1995. CYC: A large-scale investment in knowledge infrastructure. Commun ACM, 38: 33–38
Lu F, Yu L, Qiu P. 2017. On geographic knowledge graph (in Chinese). J Geo-inform Sci, 19: 723–734
Lucas S G. 2018. The GSSP method of chronostratigraphy: A critical review. Front Earth Sci, 6: 191
Ma X G, Ma C, Wang C B. 2020. A new structure for representing and tracking version information in a deep time knowledge graph. Comput Geosci, 145: 104620
Mitraka E, Waagmeester A, Su A, Good B. 2015. Wikidata: A central hub for linked open life science data. In: the Biocuration 2015 Conference. Beijing
Nakashole N, Theobald M, Weikum G. 2011. Scalable knowledge harvesting with high precision and high recall. In: Proceedings of the Fourth ACM International Conference on Web Search and Data Mining. 227–236
NSTC (National Science and Technology Council). 2018. Open Knowledge Network. 8
OECD (Organization for Economic Cooperation and Development). 1996. The Knowledge-based Economy. Paris: OECD
Oramas S, Ostuni V C, Noia T D, Serra X, Sciascio E D. 2017. Sound and music recommendation with knowledge graphs. ACM Trans Intell Syst Technol, 8: 1–21
Singhal A. 2012. Introducing the Knowledge Graph: Things, not strings. Google Blog. https://www.blog.google/products/search/introducing-knowledge-graph-things-not/
Sun H. 2017. Encyclopedia of Geoscience (in Chinese). Beijing: Science Press
Tang J. 2020. On the next decade of artificial intelligence (in Chinese). CAAI Trans Intell Syst, 15: 193–198
Tansley S, Tolle K. 2009. The Fourth Paradigm: Data-Intensive Scientific Discovery. Redmond, WA: Microsoft Research
Tay Y, Luu A T, Hui S C. 2017. Non-Parametric estimation of multiple embeddings for link prediction on dynamic knowledge graphs. In: Proceedings of the Thirty First Conference on Artificial Intelligence (AAAI). Menlo Park: AAAI. 1243–1249
Walsh S, Gradstein F, Ogg J. 2004. History, philosophy, and application of the Global Stratotype Section and Point (GSSP). Lethaia, 37: 201–218
Wang P, Jian Z. 2019. Exploring the deep South China Sea: Retrospects and prospects. Sci China Earth Sci, 62: 1473–1488
Wang S, Zhang X, Ye P, Du M, Lu Y, Xue H. 2019. Geographic Knowledge Graph (GeoKG): A formalized geographic knowledge representation. ISPRS Int J Geo-Inform, 8: 184
Wang Z, Li J. 2016. Text-Enhanced representation learning for knowledge graph. In: Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence. Menlo Park: AAAI. 1293–1299
Xiong W. 2019. Influence of artificial intelligence on the development of some fields of surveying and mapping tech (in Chinese). Geomat Inform Sci Wuhan Univ, 44: 101–105
Xu J, Pei T, Yao Y. 2010. Conceptual framework and representation of geographic knowledge map. Geo-Inf Sci, 12: 496–502
Zhai G, Yang S, Chen N. 2018. Big Data epoch: Challenges and opportunities for geology (in Chinese). Bull Chin Acad Sci, 8: 825–831
Zhang X, Zhang C, Wu M, Lv G. 2020. Spatiotemporal features based geographical knowledge graph construction. Sci Sin-Inf, 50: 1019–1032

(Responsible editor: Xin LI)