GEOINFORMATICS

Original article
https://doi.org/10.21285/2686-9993-2021-44-3-204-218

Construction and applications of knowledge graph of porphyry copper deposits

Yongzhang Zhou[a], Qianlong Zhang[b], Wenjie Shen[c], Fan Xiao[d], Yanlong Zhang[e], Shiwu Zhou[f], Yongjian Huang[g], Junjie Ji[h], Lei Tang[i], Chong Ouyang[j]
[a–d,h–j]*Sun Yat-sen University, Guangzhou, China*
[a–d,h–j]*Guangdong Provincial Key Lab of Geological Processes and Mineral Resource Survey, Guangzhou, China*
[e,f]*Guangdong Institute of High Quality Resources and Environment, Guangzhou, China*
[g]*Guangdong Xuanyuan Network Tech. Inc., Guangzhou, China*
Corresponding author: Yongzhang Zhou, zhouyz@mail.sysu.edu.cn

Abstract. A knowledge graph is becoming popular due to its ability to describe the real world by using a graph language that can be understood by both humans and machines using computer technologies. A case study to construct the knowledge graph of porphyry copper deposits is presented in this paper. First of all, the raw text data is collected and integrated from selected porphyry copper deposits and porphyry-skarn copper deposits in the Qinzhou Bay – Hangzhou Bay metallogenic belt, South China. Second, the text's entities, relations, and attributes are labeled and extracted with reference to the conceptual model of porphyry copper deposits in the study area. The third, a knowledge graph of porphyry copper deposits, was constructed using Neo4j 4.3. The resulted knowledge graph of porphyry copper deposit has the basic functions of an application. Furthermore, as part of a planned integrated knowledge graph from a single deposit, through an upper-geared metallogenic series, to a high-top metallogenic province, the understanding from the present study may be extended to mineral resource prospectivity and assessment beyond today. The interrelationship between the earth system, the metallogenic system, the exploration system, and the prospectivity and assessment (ES-MS-ES-PS) should be completely understood, and a knowledge graph system for ES-MS-ES-PS is needed. The key scientific and technological problems for achieving the ES-MS-ES-PS knowledge graph system are included in the progressively relative system of the domain ontology and knowledge graph of ES-MS-ES-PS, the automatic construction technology of complicated ES-MS-ES-PS domain ontology and knowledge graph, the self-evolution and complementary techniques for multi-modal correlation data embedding in the ES-MS-ES-PS knowledge graph, and the knowledge graph, big data mining and artificial intelligence based on ES-resource prospectivity, and assessment theory, and methods.

Keywords: geological knowledge graph, geological big data, prospectivity and assessment of mineral resource, domain ontology, porphyry copper deposit

Funding: this work was supported by the Major Project of the National Natural Science Foundation of China (U1911202); Guangdong Provincial Key R&D Project (2020B1111370001); Guangdong Provincial Science and Technology Commissioner Project (GDKTP2020053500).

For citation: Zhou Yongzhang, Zhang Qianlong, Shen Wenjie, Xiao Fan, Zhang Yanlong, Zhou Shiwu, et al. Construction and applications of knowledge graph of porphyry copper deposits. *Nauki o Zemle i nedropol'zovanie = Earth sciences and subsoil use.* 2021;44(3):204-218. https://doi.org/10.21285/2686-9993-2021-44-3-204-218.

©Zhou Yongzhang, Zhang Qianlong, Shen Wenjie, Xiao Fan, Zhang Yanlong, Zhou Shiwu, Huang Yongjian, Ji Junjie, Tang Lei, Ouyang Chong, 2021

ГЕОИНФОРМАТИКА

Оригинальная статья
УДК 550.8.053

Построение и применение графа знаний медно-порфировых месторождений

Юнчжан Чжоу[a], Цяньлун Чжан[b], Вэньцзе Шэнь[c], Фань Сяо[d], Яньлун Чжан[e], Шиу Чжоу[f], Юнцзянь Хуан[g], Цзюньцзе Цзи[h], Лэй Тан[i], Чун Оуян[j]

[a–d,h–j]*Университет им. Сунь Ятсена, г. Гуанчжоу, Китай*
[a–d,h–j]*Центральная лаборатория службы геологических процессов и минеральных ресурсов провинции Гуандун, г. Гуанчжоу, Китай*
[e,f]*Гуандунский институт высококачественных ресурсов и окружающей среды, г. Гуанчжоу, Китай*
[g]*Компания Гуандун Сюаньюань Сеть и Технологии Инкорпорейтед, г. Гуанчжоу, Китай*
Автор, ответственный за переписку: Чжоу Юнчжан, zhouyz@mail.sysu.edu.cn

Резюме. Граф знаний становится популярным благодаря своей способности описывать с использованием компьютерных технологий реальный мир при помощи языка графов, понятного как людям, так и машинам. В данной статье представлен пример построения графа знаний медно-порфировых месторождений. Во-первых, необработанные текстовые данные собраны и интегрированы по выбранным месторождениям медно-порфировых и скарново-порфировых медных месторождений в металлогеническом поясе заливов Циньчжоу – Ханчжоу Южного Китая. Во-вторых, текстовые сущности, отношения и атрибуты помечены и извлечены со ссылкой на концептуальную модель медно-порфировых месторождений в районе исследования. В-третьих, граф знаний медно-порфировых месторождений был построен с использованием Neo4j 4.3. Полученный граф знаний месторождения медно-порфировых руд имеет основные функции приложения. Кроме того, как часть запланированного интегрированного графа знаний от единичного месторождения транслируется через металлогеническую серию до крупной металлогенической провинции, поэтому результаты настоящего исследования могут быть со временем распространены на перспективность и оценку минеральных ресурсов других месторождений. Взаимосвязь между земной системой, металлогенической системой, системой разведки и оценки перспективности (ES-MS-ES-PS) должна быть полностью понята, а для этого необходима система графа знаний для ES-MS-ES-PS. Ключевые научные и технологические проблемы для создания системы графа знаний ES-MS-ES-PS включены в прогрессивную относительную систему онтологии предметной области и графа знаний ES-MS-ES-PS, технологии автоматического построения сложных онтологий предметной области MS-ES-PS и графа знаний, саморазвитие и дополнительные методы для встраивания данных многомодальной корреляции в граф знаний ES-MS-ES-PS, а также построение графа знаний, интеллектуальный анализ больших данных и искусственный интеллект на основе перспективности ресурсов земной коры, теории и методов оценки.

Ключевые слова: граф геологических знаний, большие геологические данные, перспективность и оценка минеральных ресурсов, онтология предметной области, медно-порфировое месторождение

Финансирование: данное исследование проводилось при поддержке Основного проекта Государственного фонда естественных наук Китая (U1911202), стратегического научно-исследовательского проекта провинции Гуандун (2020B1111370001), проекта уполномоченного по науке и технологиям провинции Гуандун (GDKTP2020053500).

Для цитирования: Чжоу Юнчжан, Чжан Цяньлун, Шэнь Вэньцзе, Сяо Фань, Чжан Яньлун, Чжоу Шиу [и др.]. Построение и применение графа знаний медно-порфировых месторождений // Науки о Земле и недропользование. 2021. Т. 44. № 3. С. 204–218. https://doi.org/10.21285/2686-9993-2021-44-3-204-218.

Introduction

In the present era of big data, data grows explosively, and it tends to be massive, heterogeneous, and loosely organized, bringing serious challenges to effective access to information and knowledge. The fundamental way out is to extend the human brain with the help of machine and machine learning, which urgently needs a language that people and machines can understand together [1–3].

The knowledge graph is one among these technologies. It is an integral part of artificial intelligence technology, known as interpretable artificial intelligence. With its powerful semantic processing ability and open organization ability, it provides an effective tool for knowledge organization and intelligent applications of information in the era of big data. Since the knowledge graph was formally proposed by Google in 2012, it has attracted much attention of researchers and has

been widely used in intelligent search, intelligent Q & A, personalized recommendation, and so on [4–12][1].

This paper presents a case study to construct the knowledge graph, with a porphyry copper mine as a carrier. It introduces the construction algorithm of the geological deposit domain knowledge graph and discusses the extended idea of knowledge graph to Earth system - Metallogenic system - Exploration system - Prediction and evaluation system.

Methodology

The basis of the knowledge graph is a semantic network to reveal the relationship between entities [13]. It describes the real-world things and their relations in the way of "graph" and stores them in the database in the way of "entity-relationship – entity" triple. As a network, it consists of nodes and edges. A node represents an entity: all kinds of things, existence, and concepts in the real world, which can be either a concrete entity or an abstract concept, such as a known ore point or an abstract porphyry copper concept. Edge represents the relationship between entities, which is represented as attributes in many scenes, such as the location of ore point, rock mass, element content or mineralization time, and process of an ore occurrence. Figure 1 is the representation diagram of knowledge graph entity, attribute, and relationship.

The complete knowledge graph architecture includes Knowledge acquisition, Knowledge representation, Knowledge storage, Knowledge modeling, Knowledge fusion, Knowledge Computing, Knowledge operation, and maintenance, etc. It is included among the next key technologies and processes:

Ontology modeling. The data model of the knowledge graph is established. In the ontology model, it is needed to construct the concept, attribute, and relationship of ontology. The process of ontology modeling is the basis of the knowledge graph. The high-quality data model can avoid many unnecessary and repetitive knowledge acquisitions, effectively improve the efficiency of knowledge graph and reduce the cost of domain data fusion.

Knowledge acquisition. In the real world, knowledge exists in structured, semi-structured, and unstructured data. Through knowledge extraction technology, different structures and types of data can be extracted into structured data that can be understood and calculated by computer. Knowledge acquisition is to extract knowledge from data of different sources and structures, form structured knowledge and store it in the knowledge graph. For text data, the extraction problems of knowledge acquisition include entity extraction, relationship extraction, attribute extraction, and event extraction.

Knowledge storage. The underlying storage method is designed to store all kinds of knowledge, so as to support the effective management and calculation of large-scale graph data. The objects of knowledge storage include

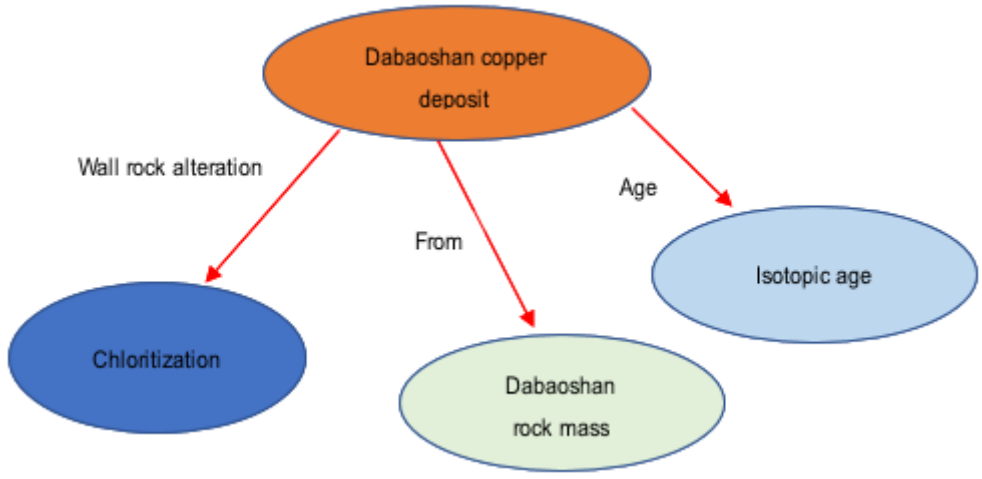

Fig. 1. Schematic diagram showing the entity, attributes, and relations of a knowledge graph
Рис. 1. Схематическая диаграмма, показывающая сущность, атрибуты и отношения графа знаний

[1] Peak labs. About OpenKG.CN. *Professional Committee of Language and Knowledge Computing Chinese Information Processing Society of China*. Available from: http://wp.openkg.cn/?page_id=77 [Accessed 28th February 2021].

basic attribute knowledge, association knowledge, event knowledge, time sequence knowledge, and resources knowledge et. The quality of knowledge storage directly affects the efficiency of knowledge query, knowledge calculation, and knowledge update in the knowledge graph.

Knowledge fusion. Knowledge fusion aims at generating new knowledge and integrating the knowledge from loosely coupled sources to form a synthetic resource to supplement incomplete knowledge and acquire new knowledge. It is an interdisciplinary subject of knowledge organization and information fusion. Hidden or valuable new knowledge can be obtained, the structure and connotation of knowledge be optimized, and knowledge services be provided, through the acquisition, matching, integration, mining, and other processing methods of knowledge on many scattered and heterogeneous resources.

Knowledge operation and maintenance. It is necessary for the real scene to iterated or evolve and improve the full knowledge graph according to the application feedback, the emerging knowledge of the same type, and the new knowledge sources after the initial construction of the knowledge graph. In the process of knowledge operation and maintenance, it is needed to ensure that the quality of the knowledge graph can be well controlled and gradually enriched. The operation and maintenance process of a knowledge graph is an engineering system, covering the whole life cycle of knowledge graph from knowledge acquisition to knowledge computing.

Usually, three basic steps are needed in the construction of a knowledge graph: (1) Information extraction, which extracts entities, attributes, and relationships among entities from unstructured and semi-structured data sources. (2) Information fusion, which eliminates the ambiguity of concepts, eliminates redundant and wrong concepts and ensures the quality of knowledge. (3) Knowledge processing, which includes quality evaluation or reasoning expansion of knowledge to obtain structured and networked knowledge system.

Structured data and text ones are the main sources of knowledge. The more commonly used tools for acquiring knowledge from structured databases are Triplify, D2RServer, OpenLink, SparqlMap, Ontop, etc. Knowledge graph visualization has Citespace, Protégé, Neo4j, and so on. Citespace is an information visualization software developed by using java language. Based on co-citation analysis theory and pathfinder algorithm, Citespace measures the literature (collections) in specific fields to find out the key path and knowledge inflection point of discipline evolution. By drawing a series of visual graphs, the potential dynamic mechanism of discipline evolution can be analyzed, and the frontier of discipline development be explored. Protégé is open-source software for ontology editing and knowledge acquisition developed by the Center for Bioinformatics, Stanford Medical School, based on Java language. Protégé is an ontology development tool and a knowledge-base editor. It is the core development tool of ontology construction in the semantic web. It provides the construction of ontology concept class, relationship, attribute and instance, and shields the specific ontology description language. Users only need to construct a domain ontology model at the conceptual level. Neo4j is a high-performance NoSQL graphic database, which stores structured data on the network. It is a high-performance graph engine, with the advantages of embedded, high performance, lightweight, and so on [14–16][2].

Knowledge graph of porphyry copper deposits

The main processes are involved in constructing the Knowledge graph of porphyry copper deposits as following:

(1) Raw data acquisition. In the Qinzhou Bay – Hangzhou Bay metallogenic belt of South China, six porphyry copper deposits and porphyry skarn-type copper deposits are selected as the experimental objects. The Dexing copper deposit, Yongping copper deposit, Qibaoshan copper deposit, Baoshan copper deposit, Dabaoshan copper deposit, and Yuanzhuding copper deposit are included among them. The relevant geological and mineral survey and published academic papers are systematically collected to form the initial data.

[2] Neo Technology, Inc. Neo4j, the world's leading graph database. *Neo4j Graph Database*. Available from: http://neo4j.com/ [Accessed 28th February 2021].

(2) The initial data acquisition, and the entity, relationship, and attribute annotation and extraction based on the conceptual model of porphyry copper deposit. The Xuanyuan data annotation system is used for data annotation. The annotation system is a general annotation system based on GUI, which allows the annotation file to be divided into multiple annotation tasks and allows multiple users to annotate and review. The system provides data annotation services, including batch storage and management of annotation files, auxiliary tools to simplify the difficulty of manual annotation, and machine annotation for specific fields.

The data extracted from the text is classified and standardized into three tuple formats with five columns: entity, entity type, relationship, attribute and attribute type. The entity is an existence of a deposit, an actual deposit, such as the Dexing porphyry copper deposit. The entity type is the type of deposit, such as porphyry copper deposit. Attribute is the attribute of deposit, which is used to describe the characteristics of deposit. Attribute type is the type of attribute. A section of the standardized CSV data is shown in Table.

(3) Graph generation. Python is used to read data and write them into the Neo4j graph database. Create a new local database in neo4j, name the database, and then import the existing data in CSV format into the py2neo database to generate a knowledge graph (Fig. 2).

The knowledge graph resulted from this case has the basic application function of a normal knowledge graph. In Neo4j, Cypher statements can be used to query the whole database, specific label query, shortest path query, where predicate query, keyword query, relational query, attribute addition and deletion, label addition and deletion, etc.

Query a label, for example, that has a directed relationship with a node. Input: match (a:`Porphyry copper deposit`{name:'Dexing copper deposit'}) – (b) return a, b, the nodes connected with the Dexing Copper Mine are gotten (Fig. 3), from which the geological information and metallogenic conditions related to the formation of Dexing copper mine are demonstrated.

Prospect: Knowledge graph of ES-MS-ES-PS

The case above is part of the ongoing experiment to build knowledge graph series from single ore deposit, through metallogenic series, to metallogenic province, aiming at providing a demo for future large-scale construction and application of knowledge graph of ore deposits.

It is reasonable to build the knowledge graph of porphyry copper deposit for the first since its metallogenic model is classic and well recognized by almost all geologists. It is also the main theoretical model for prospecting for porphyry copper deposits. The workload of building an ontology model is relatively controllable. Based on the existing geological survey reports and other unstructured and semi-structured data, through ontology construction, knowledge extraction, knowledge disambiguation, and knowledge fusion, the knowledge graph of porphyry copper deposit may well be constructed.Similarly, the knowledge graph of epithermal metallogenic system (Fig. 4) and Qinzhou Bay – Hangzhou Bay metallogenic belt (Fig. 5) can be constructed.

Individual deposits, metallogenic series, and important metallogenic areas (belts) contain

Standardized data (part)
Стандартизованные данные (часть)

Entity	Entity type	Relationship	Attribute	Attribute type
Yuanzhuding deposit	Porphyry Copper Deposit	Magmatism	Diorite granite	Types of magmatism
Dexing deposit	Porphyry Copper Deposit	Metamorphism	Phyllite	Types of metamorphism
Yongping deposit	Porphyry Copper Deposit	Wall rock alteration	Silicification	Types of wall rock alteration
Qibaoshan deposit	Porphyry Copper Deposit	Stratigraphic evolution	Carboniferous series	Stratigraphic type
Dabaoshan deposit	Porphyry Copper Deposit	Element enrichment and depletion	Cu	Types of element enrichment and depletion

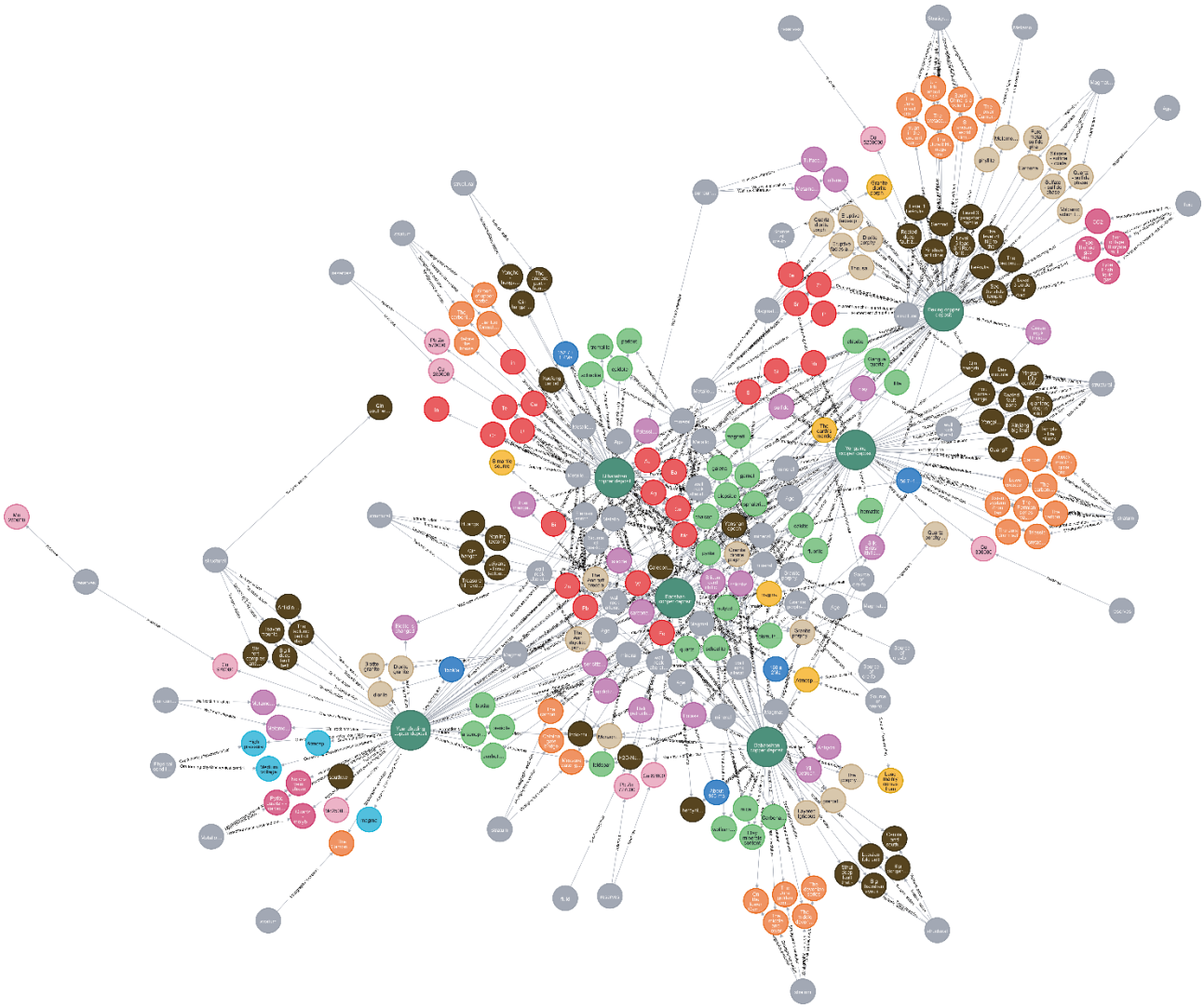

Fig. 2. Knowledge graph of porphyry copper deposit from Qinzhou Bay – Hangzhou Bay metallogenic belt, South China
Рис. 2. Граф знаний медно-порфировых отложений металлогенического пояса заливов Циньчжоу – Ханчжоу, Южный Китай

typical relationships of deposits at different levels. The single deposit belongs to the metallogenic series and the important metallogenic area (belt), and its attributes are inherited. The metallogenic series and the important metallogenic area (belt) are intersecting, and their attribute relationship is complex. The construction of a knowledge graph system of individual deposits, metallogenic series, and important metallogenic areas (belts) can provide valuable support for the construction of a larger knowledge graph of Earth system - Metallogenic system - Exploration system - Prediction and evaluation system.

The prediction and evaluation of mineral resources is one of the important directions in the application of geological science and has formed a unique theory and method system [17, 18]. But generally speaking, the existing metallogenic prediction theories and methods are mainly composed of two parts. The first is the mineral prediction model, which is the metallogenic prediction elements and criteria established by summarizing the metallogenic law of typical deposits and geophysical, geochemical, and remote sensing anomaly characteristics. The second is the mathematical model of prospecting information extraction and fusion, that is, the mathematical model is used to quantify and fuse the corresponding prediction elements in the prediction model, so as to finally estimate the size of metallogenic potential. Most of the research focuses on the mathematical model of prospecting information extraction and fusion. The research on the mineral prediction model mainly depends on the knowledge

Fig. 3. Directed relation nodes of the Dexing porphyry copper deposit
Different colors represent different types of geological properties
Рис. 3. Узлы направленной связи медно-порфирового месторождения Дексинг
Разные цвета представляют разные типы геологических свойств

drive of geological experts, i.e., a quantitative mineral prediction model is formed based on the main metallogenic geological characteristics and prospecting indicators (including geology, geophysics, geochemistry, and remote sensing) of several typical deposits.The traditional approach is flawed. First of all, its starting point is based on the characteristics of metallogenic geology and metallogenic law, that is, the characteristics of metallogenic system itself, without considering the correlation between metallogenic system and other earth systems such as disaster system and climate system, which may lead to the omission of some important prediction elements and the incompleteness of prediction model. Secondly, over reliance on the knowledge driving of geological experts reduces the effectiveness of the prediction model.

The future mineral resources prediction and evaluation should fully understand the relationship among the Earth system, the metallogenic system, the exploration system and the prediction and evaluation system (ES-MS-ES-PS). It is an important development direction to establish the associated knowledge graph system of “ES-MS-ES-PS”.

The key scientific and technical problems following need to be solved in order to establish the knowledge graph system of ‘ES-MS-ES-PS’:

(1) Progressive correlation system of the ES-MS-ES-PS knowledge graphs. The knowledge graph progressive correlation system of the Earth system - Metallogenic system - Exploration system - Prediction and evaluation system is not well understood yet. Behind them are the Earth, Metallogenic, Exploration, and Mining Sciences,

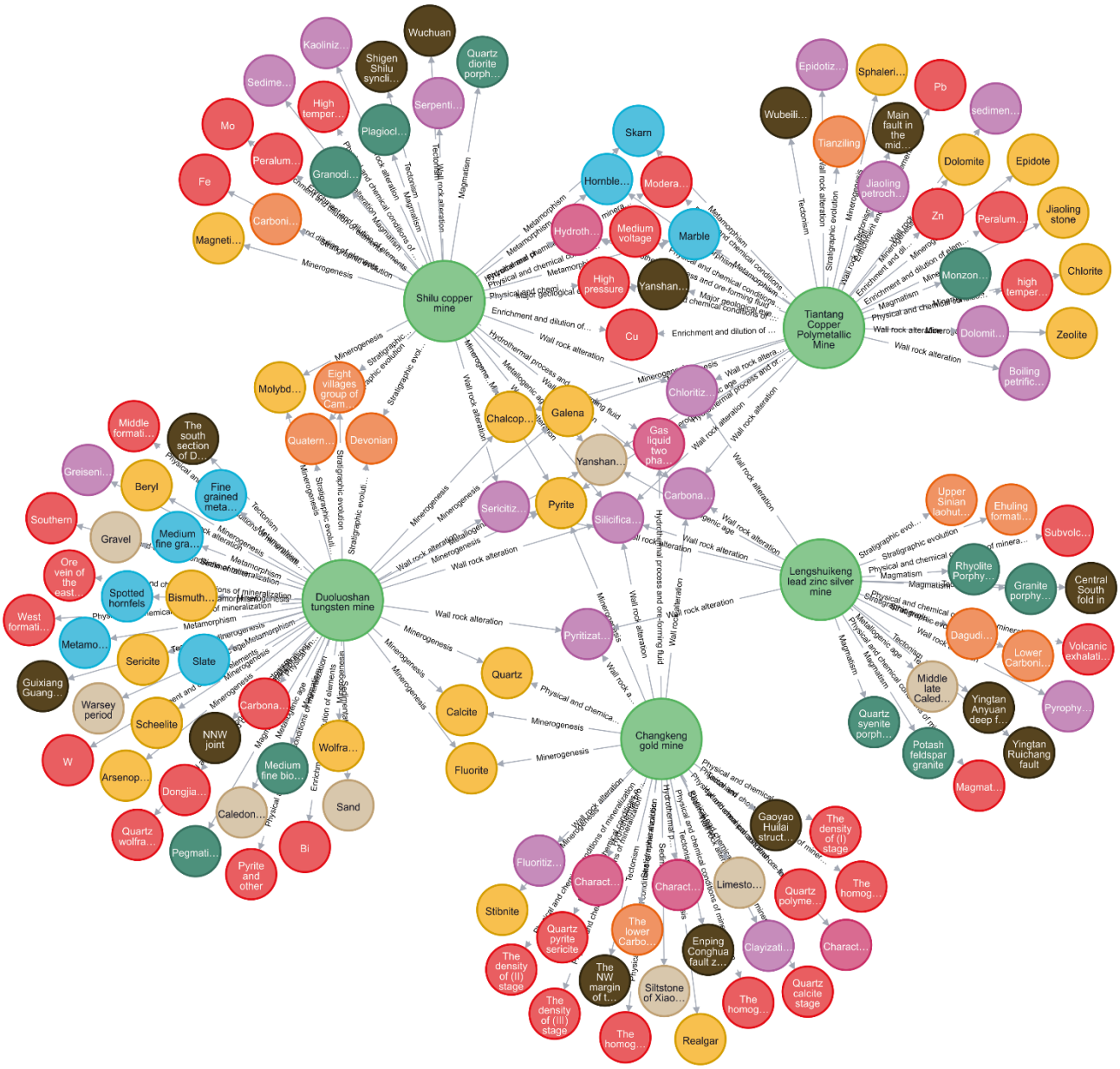

Fig. 4. The visual interface of knowledge graph of epithermal metallogenic system
Рис. 4. Визуальный интерфейс графа знаний эпитермальной металлогенической системы

which are both systematic and intricate. This limits the integration of data and knowledge, and also the exploration of the potential of the system. Based on the system association framework of the knowledge graph, the interpretable prediction and evaluation of mineral resources can be formed through the digestion and fusion of knowledge co-index and the community detection and correlation based on graph theory.

The ES-MS-ES-PS can be regarded as self-contained but interrelated systems. Logically, the earth system includes the metallogenic system, which inherits the attributes and relations of the earth system. The earth system has a larger extension, and the metallogenic system has a more specific connotation. The exploration system and the prediction and evaluation system are the current expert knowledge systems. They are not completely coincident with the actual metallogenic system, and there is an intersection between them.

(2) The ontology construction of ES-MS-ES-PS. Geological big data are considered as the main research object. Firstly, a machine learning algorithm is used to model and associate the knowledge of the Earth system, metallogenic system, exploration system, and prediction and evaluation system under the guidance of the ontology model of ES-MS-ES-PS. Speech tagging has done for text data. Then, the candidate entity

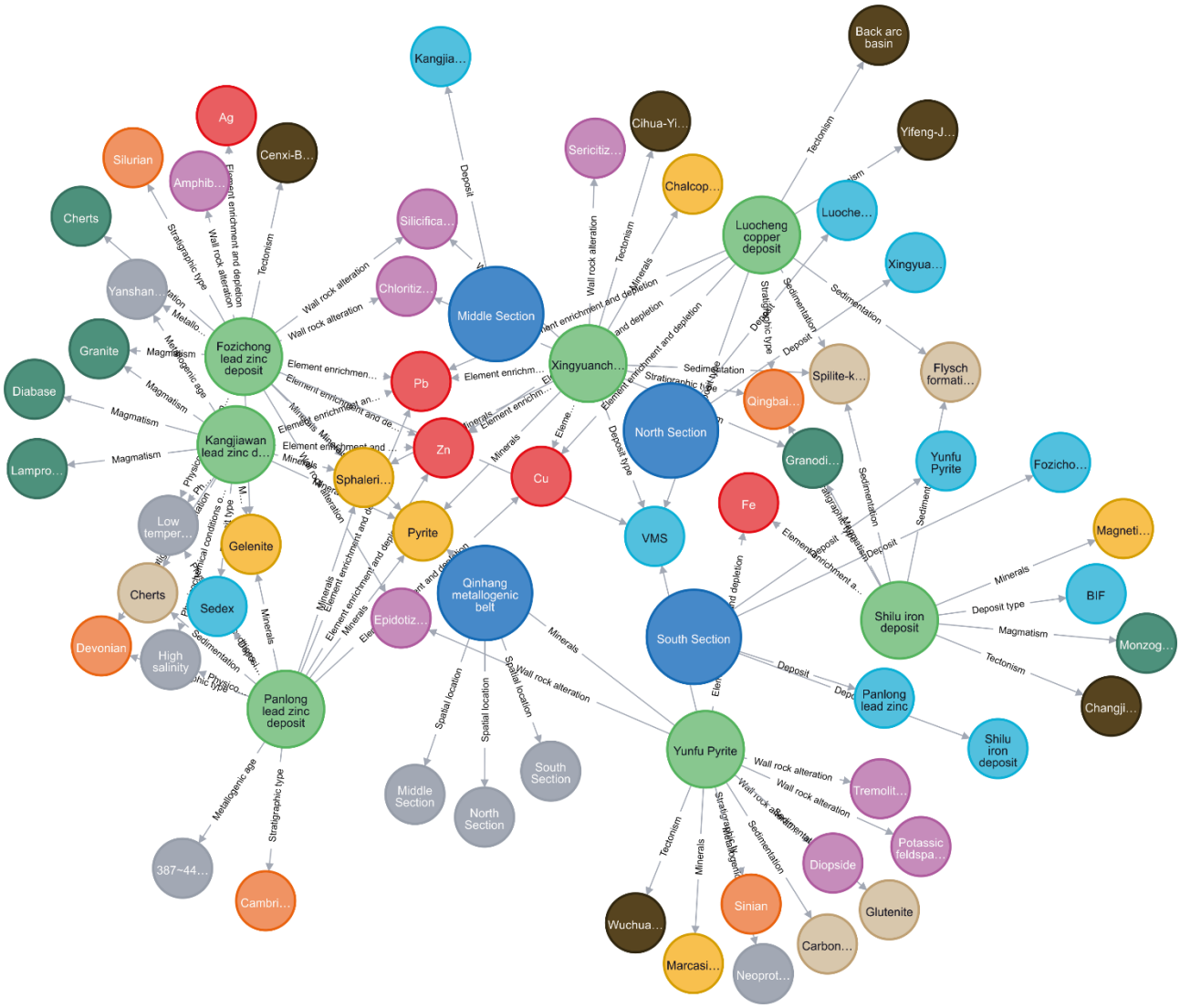

Fig. 5. The visual section of knowledge graph of the Qinzhou Bay – Hangzhou Bay metallogenic belt
Рис. 5. Визуальный разрез графа знаний металлогенического пояса заливов Циньчжоу – Ханчжоу

pairs and relationship features are extracted, and the factor graph model method is used to train the extraction rules, which are used to extract the domain entities and semantic relationships of the ES-MS-ES-PS according to the results of part of speech tagging. Finally, the extracted entity and semantic relationship are stored in the form of the graph database, and the knowledge base of ES-MS-ES-PS is established and visualized. The corresponding knowledge graph and data sharing platform of ES-MS-ES-PS are established in order to realize the information retrieval, acquisition, sharing, and logical reasoning of the knowledge base of ES-MS-ES-PS.

Furthermore, taking the geological ontology of ore controlling elements in the fields of ES-MS-ES-PS as a bridge, the ES-MS-ES-PS are organically linked. Furthermore, taking the geological ontology of ore controlling elements in the fields of ES-MS-ES-PS as a bridge, the ES-MS-ES-PS are organically linked.

Through machine learning, semantic analysis, visual analysis, and other intelligent methods, the ES-MS-ES-PS are analyzed. The in-depth development of knowledge graphs in the field of exploration systems and prediction and evaluation systems provides multi-source, multi-dimensional, spatiotemporal, multi-scale information and knowledge intelligent services for mineral resource prediction and evaluation, improves the breadth, accuracy, and efficiency of deep-sitting prospecting information identification and extraction, and links and integrates prospecting information in the fields of ES-MS-ES-PS. All above will lead to the occurrence of the smart prediction of mineral resources based on ES-MS-ES-P.

(3) Automatic extraction technology of large-scale geological knowledge graph relationship. In the process of automatic acquisition of geological knowledge and construction of knowledge graph, relation extraction is the core and the only way to accomplish this task. The purpose of relation extraction is to extract the relationship between entities from unlabeled self-owned texts, and then structure the entity and relationship into structured knowledge, and extend it into a knowledge graph accordingly. The traditional relational extraction method is based on the construction of a supervised extraction system, and its training and deployment rely heavily on large-scale manually labeled data, which consumes huge time and manpower. This project develops and constructs a remote supervised relation extraction system to make up for the problems existing in the traditional supervised model. At the same time, it explores the introduction of multi-source external information to eliminate the noise problem in remote supervision and alleviate the impact of long-tail data, so as to obtain a more robust geological knowledge extraction system.

(4) Evolving and improving itself of knowledge graph embedding multi-modal association data. Heterogeneity is an unneglectable problem to construct an opening geological knowledge graph. The traditional way to solve ontology heterogeneity is ontology integration. Ontology integration directly merges multiple ontologies into a large ontology, and each heterogeneous system uses the unified ontology. In this way, the interaction between them can be carried out directly, thus solving the problem of ontology heterogeneity. However, the integration of ontology is time-consuming and laborious and lacks automatic method support. With the change of multiple ontologies, the integration process needs to be repeated and the cost is too high. In addition, the integrated ontology is not universal and flexible for different applications. Therefore, ontology integration is not suitable to solve the distributed and dynamic multi ontology application problems in the knowledge graph. In fact, most applications only need to realize the interoperability between ontologies to meet the requirements, and complete integration is not necessary. This project studies the ontology mapping method based on multi-modal association data embedding. It achieves ontology interoperability by establishing mapping rules between ontologies. At the same time, it introduces a large number of texts, images, and numerical information in the knowledge base, improves the quality of mapping and matching, and realizes the effective completion of the knowledge graph.

(5) Data acquisition, access, and fusion mechanism based on the knowledge graph. Community structure is popular in the geological knowledge graph. Community refers to a group of nodes that are closely related to each other within the community, and their relationship with nodes outside the community is relatively loose. It has many applications to obtain and query community data, identify community structure, analyze the structure and function of the whole network, and predict the interaction between various elements of the network, such as geological network analysis, identification of special geological phenomena, deposit prediction, etc. Traditional community detection only considers the structural features with neglecting the necessary semantic information on the knowledge graph. This project will study the community detection algorithm for knowledge graphs, and introduce attribute-based retrieval, which can effectively improve computational efficiency.

(6) The construction norms and standard system of the geoscience knowledge graph. The standardization of geological knowledge graph is greatly important to improve construction efficiency, ensure data re-use in multiple fields, and give full play to knowledge graph analysis and technical value. This project studies the overall framework of the geological knowledge graph, mainly focusing on knowledge acquisition, knowledge representation, knowledge modeling, knowledge fusion, knowledge storage, knowledge computing, knowledge operation and maintenance, natural language processing, and other related supporting technology fusion, covering the whole life cycle of the knowledge graph, providing guarantee for technology development and application.

Conclusions

It may be concluded through the analysis above that:

(1) Knowledge graph represents the objects and their relationships in the objective world with the mathematical model of the graph, which makes knowledge and data easier to exchange, circulate, and process between computers and between computers and people. Compared with a traditional relational database, a knowledge graph is more flexible and more suitable for a big data environment. In the era of big data and artificial intelligence, there is an urgent need for a language that people and machines can understand together to extend the human brain.

(2) The construction of the knowledge graph of porphyry copper deposits is a good experiment, it may be well extended to the epithermal metallogenic system and the Qinzhou Bay – Hangzhou Bay, metallogenic belt, South China, resulting in a complete knowledge graph system from the single deposit, through metallogenic series, to an important metallogenic area (belt). Then a greater knowledge graph system of Earth system - Metallogenic system - Exploration system - Prediction and evaluation system may be expected.

(3) The future mineral resource prediction and evaluation should fully understand the relationship among the Earth system, the metallogenic system, the exploration system, and the prediction and evaluation system. A more universal metallogenic prediction system may be established through open integration and deep mining of different systems or geological big data.

The transformation of quantitative prediction and evaluation of mineral resources may be promoted by the establishment of the associated knowledge graph system of the Earth system - Metallogenic system - Exploration system - Prediction and evaluation system.

References

1. Zhang Q., Zhou Y. Big data helps geology develop rapidly. *Acta Petrologica Sinica*. 2018;34(11):3167-3172. (In Chinese).

2. Zhou Y., Wang J., Zuo R., Xiao F., Shen W., Wang S. Machine learning, deep learning and Python language. *Acta Petrologica Sinica.* 2018;34(11):3173-3178. (In Chinese).

3. Zhou Y., Zhang L., Zhang O., Wang J. *Big data mining & machine learning in geoscience*. GuangZhou: Sun Yat-sen University Press; 2018. 269 p. (In Chinese).

4. Singhal A. Introducing the Knowledge Graph: things, not strings. *Blog.google*. Available from: https://www.blog.google/products/search/introducing-knowledge-graph-things-not [Accessed 28th February 2021].

5. Wu W., Li H., Wang H., Zhu K. Q. Probase: a probabilistic taxonomy for text understanding. *Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data*. 2012:481-492. https://doi.org/10.1145/2213836.2213891.

6. Hoffart J., Suchanek F. M., Berberich K., Weikum G. YAGO2: a spatially and temporally enhanced knowledge base from Wikipedia. *Artificial Intelligence*. 2013;194:28-61. https://doi.org/10.1016/j.artint.2012.06.001.

7. Lukovnikov D., Fischer A., Lehmann J., Auer S. Neural network-based question answering over knowledge graphs on word and character level. *WWW'17: Proceedings of the 26th International Conference on World Wide Web*. 2017:1211-1220. https://doi.org/10.1145/3038912.3052675.

8. Xu B., Xu Y., Liang J., Xie C., Liang B., Cui W., et al. CN-DBpedia: a never-ending Chinese Knowledge extraction system. *Advances in Artificial Intelligence: From Theory to Practice*. 2017:428-438. https://doi.org/10.1007/978-3-319-60045-1_44.

9. Palumbo E., Rizzo G., Troncy R., Baralis E., Osella M., Ferro E. An empirical comparison of knowledge graph embeddings for item recommendation. *Istituzionale della Ricerca*. 2018. Available from: https://iris.polito.it/retrieve/handle/11583/2710124/203256/paper2.pdf [Accessed 28th February 2021].

10. Wang C., Yu H., Wan F. Information retrieval technology based on knowledge graph. *Proceedings of the 2018 3rd International Conference on Advances in Materials, Mechatronics and Civil Engineering (ICAMMCE 2018)*. 2018. https://doi.org/10.2991/icammce-18.2018.65.

11. Qi H., Dong S., Zhang L., Hu H., Fan J. Construction of Earth science knowledge graph and its future perspectives. *Geological Journal of China Universities.* 2020;26(1):2-10. (In Chinese). https://doi.org/10.16108/j.issn1006-7493.2019099.

12. Zhou Y., Zhang Q., Huang Y., Yang W., Xiao F. Construction of knowledge graph of porphyry copper deposit from Qingzhou Bay – Hangzhou Bay and insight into knowledge graph based mineral resource prediction and evaluation. *Earth Sciences Frontiers*. 2021;28(3):67-75. (In Chinese).

13. Liu Q., Li Y, Duan H, Liu Y, Qin Z. Knowledge graph construction techniques. *Journal of Computer Research and Development.* 2016;53(3):582-600. (In Chinese). https://doi.org/10.7544/issn1000-1239.2016.20148228.

14. Sahoo S., Halb W., Hellmann S., Idehen K., Thibodeau Jr T., Auer S., et al. A survey of current approaches for mapping of relational databases to RDF: W3C RDB2RDF Incubator Group report. *W3.org*. Available from: https://www.w3.org/2005/Incubator/rdb2rdf/RDB2RDF_SurveyReport.pdf [Accessed 28th February 2021].

15. Chen Y., Chen C., Liu Z., Hu Z., Wang X. The methodology function of CiteSpace mapping knowledge domains. *Studies in Science of Science*. 2015(2):243-252. (In Chinese). https://doi.org/10.16192/j.cnki.1003-2053.2015.02.009.

16. Knublauch H., Fergerson R. W., Noy N. F., Musen M. A. The Protégé OWL plugin: an open development environment for semantic web applications. *The Semantic Web – ISWC 2004*. 2004:229-243. https://doi.org/10.1007/978-3-540-30475-3_17.

17. Zhao P. Quantitative mineral prediction and deep mineral exploration. *Earth Science Frontiers*. 2007;14(5):1-10. (In Chinese).

18. Agterberg F. *Geomathematics: theoretical foundations, applications and future developments*. Springer International Publishing; 2014. 553 p. Available from: https://www.springer.com/gp/book/9783319068732 [Accessed 28th February 2021].

Список источников

1. Zhang Q., Zhou Y. Big data helps geology develop rapidly // Acta Petrologica Sinica. 2018. Vol. 34. Iss. 11. P. 3167–3172.

2. Zhou Y., Wang J., Zuo R., Xiao F., Shen W., Wang S. Machine learning, deep learning and Python language // Acta Petrologica Sinica. 2018. Vol. 34. Iss. 11. P. 3173–3178.

3. Zhou Y., Zhang L., Zhang O., Wang J. Big data mining & machine learning in geoscience. GuangZhou: Sun Yat-sen University Press, 2018. 269 p.

4. Singhal A. Introducing the Knowledge Graph: things, not strings // Blog.google [Электронный ресурс]. URL: https://www.blog.google/products/search/introducing-knowledge-graph-things-not/ (28.02.2021).

5. Wu W., Li H., Wang H., Zhu K. Q. Probase: a probabilistic taxonomy for text understanding // Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data. 2012. P. 481–492. https://doi.org/10.1145/2213836.2213891.

6. Hoffart J., Suchanek F. M., Berberich K., Weikum G. YAGO2: a spatially and temporally enhanced knowledge base from Wikipedia // Artificial Intelligence. 2013. Vol. 194. P. 28–61. https://doi.org/10.1016/j.artint.2012.06.001.

7. Lukovnikov D., Fischer A., Lehmann J., Auer S. Neural network-based question answering over knowledge graphs on word and character level // WWW'17: Proceedings of the 26th International Conference on World Wide Web. 2017. P. 1211–1220. https://doi.org/10.1145/3038912.3052675.

8. Xu B., Xu Y., Liang J., Xie C., Liang B., Cui W., et al. CN-DBpedia: a never-ending Chinese Knowledge extraction system // Advances in Artificial Intelligence: From Theory to Practice. 2017. P. 428–438. https://doi.org/10.1007/978-3-319-60045-1_44.

9. Palumbo E., Rizzo G., Troncy R., Baralis E., Osella M., Ferro E. An empirical comparison of knowledge graph embeddings for item recommendation // Istituzionale della Ricerca. 2018. [Электронный ресурс]. URL: https://iris.polito.it/retrieve/handle/11583/2710124/203256/paper2.pdf (28.02.2021).

10. Wang C., Yu H., Wan F. Information retrieval technology based on knowledge graph // Proceedings of the 2018 3rd International Conference on Advances in Materials, Mechatronics and Civil Engineering (ICAMMCE 2018). 2018. https://doi.org/10.2991/icammce-18.2018.65.

11. Qi H., Dong S., Zhang L., Hu H., Fan J. Construction of Earth science knowledge graph and its future perspectives // Geological Journal of China Universities. 2020. Vol. 26. Iss. 1. P. 2–10. https://doi.org/10.16108/j.issn1006-7493.2019099.

12. Zhou Y., Zhang Q., Huang Y., Yang W., Xiao F. Construction of knowledge graph of porphyry copper deposit from Qingzhou Bay – Hangzhou Bay and insight into knowledge graph based mineral resource prediction and evaluation // Earth Sciences Frontiers. 2021. Vol. 28. Iss. 3. P. 67–75.

13. Liu Q., Li Y., Duan H., Liu Y., Qin Z. Knowledge graph construction techniques // Journal of Computer Research and Development. 2016. Vol. 53. Iss. 3. P. 582–600. https://doi.org/10.7544/issn1000-1239.2016.20148228.

14. Sahoo S., Halb W., Hellmann S., Idehen K., Thibodeau Jr T., Auer S., et al. A survey of current approaches for mapping of relational databases to RDF: W3C RDB2RDF Incubator Group report // W3.org [Электронный ресурс]. URL: https://www.w3.org/2005/Incubator/rdb2rdf/RDB2RDF_SurveyReport.pdf (28.02.2021).

15. Chen Y., Chen C., Liu Z., Hu Z., Wang X. The methodology function of CiteSpace mapping knowledge domains // Studies in Science of Science. 2015. Vol. 2. P. 243–252. https://doi.org/10.16192/j.cnki.1003-2053.2015.02.009.

16. Knublauch H., Fergerson R.W., Noy N.F., Musen M.A. The Protégé OWL plugin: an open development environment for semantic web applications // The Semantic Web – ISWC 2004. 2004. P. 229–243. https://doi.org/10.1007/978-3-540-30475-3_17.

17. Zhao P. Quantitative mineral prediction and deep mineral exploration // Earth Science Frontiers. 2007. Vol. 14. Iss. 5. P. 1–10.

18. Agterberg F. Geomathematics: theoretical foundations, applications and future developments. Springer International Publishing, 2014. 553 p. [Электронный ресурс]. URL: https://www.springer.com/gp/book/9783319068732 (28.02.2021).

Information about the authors / Информация об авторах

Yongzhang Zhou, Professor & Director of the Center for Earth Environment & Resources of Sun Yat-sent University. He got his B. Sc. degree from Sun Yat-sent University (1982), M. Sc. from The Chinese Academy of Sciences (1987), Ph. D. from Québec Université, Canada (1992), and went to Stanford University as a visiting professor cooperating with Prof. John Harbaugh in 1996. He is the winner of the Felix Chayes Prize of the International Association for Mathematical Geosciences (2015) and the Excellent Teacher of National Education Ministry of China. He serves as the Chair of Big data and Mathematical Committee of China Society for Mineralogy, Petrolgy & Geochenmistry, the Co-Chairman of the Topical Section of the IAMG for Chinese Members (IAMG-CN), the Chief-Advisor of IAMG Student Chapter at Sun Yat-sen University (IAMG-SYSU). Included are in his research interests: big data mining, machine learning and mathematical geoscience; ore deposit-related or Environmental geochemistry.

Юнчжан Чжоу – профессор и директор Центр изучения окружающей среды и ресурсов университета им. Сунь Ятсена. Он получил степень бакалавра в университете Сунь Ятсена в 1982 году, степень магистра в Китайской академии наук в 1987-м, степень доктора философии в Квебекском университете Канады в 1992-м, работал в Стэнфордском университете в Калифорнии, США, в 1996 году в качестве приглашенного профессора, сотрудничая с профессором Джоном Харбо. Юнчжан Чжоу является лауреатом премии Феликса Чейса Международной ассоциации математических наук о Земле (2015) и заслуженным профессором Министерства национального образования Китая, председателем Комитета по большим данным и математике Китайского общества минералогии, нефти и геохимии, сопредседателем тематической секции IAMG (International Association for Mathematical Geosciences) для китайских членов (IAMG-CN), главным советником студенческого отделения IAMG в Университете им. Сунь Ятсена (IAMG-SYSU). В круг его научных интересов входят: интеллектуальный анализ больших данных, машинное обучение и математические науки о Земле; геохимия рудных месторождений и экологическая геохимия.

Yongzhang Zhou,
Dr. Sci. (Geol. & Mineral.), Professor,
School of Earth Sciences & Geological Engineering,
Center for Earth Environment & Resources,
Sun Yat-sen University,
Guangzhou, China,
Guangdong Provincial Key Lab of Geological Processes and Mineral Resource Survey,
Guangzhou, China,
zhouyz@mail.sysu.edu.cn,
https://orcid.org/000-0002-8572-5849.

Чжоу Юнчжан,
доктор геолого-минералогических наук, профессор,
Школа геологических наук и геологического инжиниринга,
Центр изучения окружающей среды и ресурсов,
Университет им. Сунь Ятсена,
г. Гуанчжоу, Китай,
Центральная лаборатория службы геологических процессов и минеральных ресурсов провинции Гуандун,
г. Гуанчжоу, Китай,
zhouyz@mail.sysu.edu.cn,
https://orcid.org/000-0002-8572-5849.

Qianlong Zhang,
School of Earth Sciences & Geological Engineering,
Center for Earth Environment & Resources,
Sun Yat-sen University,
Guangzhou, China,
Guangdong Provincial Key Lab of Geological Processes and Mineral Resource Survey,
Guangzhou, China.

Чжан Цяньлун,
Школа геологических наук и геологического инжиниринга,
Центр изучения окружающей среды и ресурсов,
Университет им. Сунь Ятсена,
г. Гуанчжоу, Китай,
Центральная лаборатория службы геологических процессов и минеральных ресурсов провинции Гуандун,
г. Гуанчжоу, Китай.

Wenjie Shen,
School of Earth Sciences & Geological Engineering,
Center for Earth Environment & Resources,
Sun Yat-sen University,
Guangzhou, China,
Guangdong Provincial Key Lab of Geological Processes and Mineral Resource Survey,
Guangzhou, China.
Шэнь Вэньцзе,
Школа геологических наук и геологического инжиниринга,
Центр изучения окружающей среды и ресурсов,
Университет им. Сунь Ятсена,
г. Гуанчжоу, Китай,
Центральная лаборатория службы геологических процессов и минеральных ресурсов провинции Гуандун,
г. Гуанчжоу, Китай.

Fan Xiao,
School of Earth Sciences & Geological Engineering,
Center for Earth Environment & Resources,
Sun Yat-sen University,
Guangzhou, China,
Guangdong Provincial Key Lab of Geological Processes and Mineral Resource Survey,
Guangzhou, China.
Сяо Фань,
Школа геологических наук и геологического инжиниринга,
Центр изучения окружающей среды и ресурсов,
Университет им. Сунь Ятсена,
г. Гуанчжоу, Китай,
Центральная лаборатория службы геологических процессов и минеральных ресурсов провинции Гуандун,
г. Гуанчжоу, Китай.

Yanlong Zhang,
Guangdong Institute of High Quality Resources and Environment,
Guangzhou, China.
Чжан Яньлун,
Гуандунский институт высококачественных ресурсов и окружающей среды,
г. Гуанчжоу, Китай.

Shiwu Zhou,
Guangdong Institute of High Quality Resources and Environment,
Guangzhou, China.
Чжоу Шиу,
Гуандунский институт высококачественных ресурсов и окружающей среды,
г. Гуанчжоу, Китай.

Yongjian Huang,
Guangdong Xuanyuan Network Tech. Inc.,
Guangzhou, China.
Хуан Юнцзянь,
Компания Гуандун Сюаньюань Сеть и Технологии Инкорпорейтед,
г. Гуанчжоу, Китай.

Junjie Ji,
School of Earth Sciences & Geological Engineering,
Center for Earth Environment & Resources,
Sun Yat-sen University,
Guangzhou, China,
Guangdong Provincial Key Lab of Geological Processes and Mineral Resource Survey,
Guangzhou, China.

Цзи Цзюньцзе,
Школа геологических наук и геологического инжиниринга,
Центр изучения окружающей среды и ресурсов,
Университет им. Сунь Ятсена,
г. Гуанчжоу, Китай,
Центральная лаборатория службы геологических процессов и минеральных ресурсов провинции Гуандун,
г. Гуанчжоу, Китай.

Lei Tang,
School of Earth Sciences & Geological Engineering,
Center for Earth Environment & Resources,
Sun Yat-sen University,
Guangzhou, China,
Guangdong Provincial Key Lab of Geological Processes and Mineral Resource Survey,
Guangzhou, China.
Тан Лэй,
Школа геологических наук и геологического инжиниринга,
Центр изучения окружающей среды и ресурсов,
Университет им. Сунь Ятсена,
г. Гуанчжоу, Китай,
Центральная лаборатория службы геологических процессов и минеральных ресурсов провинции Гуандун,
г. Гуанчжоу, Китай.

Chong Ouyang,
School of Earth Sciences & Geological Engineering,
Center for Earth Environment & Resources,
Sun Yat-sen University,
Guangzhou, China,
Guangdong Provincial Key Lab of Geological Processes and Mineral Resource Survey,
Guangzhou, China.
Оуян Чун,
Школа геологических наук и геологического инжиниринга,
Центр изучения окружающей среды и ресурсов,
Университет им. Сунь Ятсена,
г. Гуанчжоу, Китай,
Центральная лаборатория службы геологических процессов и минеральных ресурсов провинции Гуандун,
г. Гуанчжоу, Китай.

Contribution of the authors / Вклад авторов

The authors contributed equally to this article.
Все авторы сделали эквивалентный вклад в подготовку публикации.

Conflict of interests / Конфликт интересов

The authors declare no conflicts of interests.
Авторы заявляют об отсутствии конфликта интересов.

The final manuscript has been read and approved by all the co-authors.
Все авторы прочитали и одобрили окончательный вариант рукописи.

Information about the article / Информация о статье

The article was submitted 03.06.2021; approved after reviewing 08.07.2021; accepted for publication 10.08.2021.
Статья поступила в редакцию 03.06.2021; одобрена после рецензирования 08.07.2021; принята к публикации 10.08.2021.