ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/361879707

Construction and application of the knowledge graph method in management of soil pollution in contaminated sites: A case study in South China

Article *in* Journal of Environmental Management · October 2022

DOI: 10.1016/j.jenvman.2022.115685

CITATIONS
15

READS
183

8 authors, including:

Feng Han
Guangxi University for Nationalities
6 PUBLICATIONS **54** CITATIONS
SEE PROFILE

Yirong Deng
Tianjin Medical University
78 PUBLICATIONS **1,021** CITATIONS
SEE PROFILE

Jun Wang
Sun Yat-sen University
18 PUBLICATIONS **249** CITATIONS
SEE PROFILE

All content following this page was uploaded by Feng Han on 07 December 2022.

The user has requested enhancement of the downloaded file.

Journal of Environmental Management 319 (2022) 115685

Contents lists available at ScienceDirect

Journal of Environmental Management

journal homepage: www.elsevier.com/locate/jenvman

Research article

Construction and application of the knowledge graph method in management of soil pollution in contaminated sites: A case study in South China

Feng Han [a,d], Yirong Deng [b], Qiyuan Liu [a,c], Yongzhang Zhou [a,d,*], Jun Wang [b], Yongjian Huang [e], Qianlong Zhang [a,d], Jing Bian [f]

[a] School of Earth Science and Engineering, Sun Yat-sen University, Zhuhai 519082, China
[b] Guangdong Provincial Key Lab of Geodynamics and Geohazards, Environmental Academy of Guangdong, Guangzhou, 510045, China
[c] Chinese Research Academy of Environment Science, Beijing, 100012, China
[d] Guangdong Provincial Key Laboratory of Geological Processes and Mineral Resources Exploration, Zhuhai, 519082, China
[e] Guangzhou Xuanyuan Research Institute, Guangzhou, 510006, China
[f] School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou, 510006, China

ARTICLE INFO

Keywords:
Knowledge graph
Urban soil pollution
Contaminated site
Environmental management

ABSTRACT

Contaminated sites are a main cause of urban soil problems and have led to increasing pollution and public risk in China as a result of the rapid growth of industrial and urban land use. Because land pollution involves extensive multi-source heterogeneous information, identifying the risk of urban soil pollution efficiently and predicting pollution-related events are important for urban environmental management. Knowledge graphs (KGs) have unique advantages in dealing with massive amounts of information. This study attempts to construct a KG of contaminated sites in South China to explore its feasibility and effectiveness in urban soil environmental management. The results demonstrate that KGs have a favorable effect in information retrieval, knowledge reasoning, and visualization. Studied cases in this article demonstrate that the KG model can achieve many functions, including the display of global information of polluted sites, and discovery of regional distribution of characteristic pollutants and main pollutants of specific industries, based on special query syntax. However, this approach is limited by some technical difficulties, such as knowledge mining of natural resources, which must be overcome in future studies to improve the operability of KG technologies.

1. Introduction

Due to the rapid socioeconomic development in China, a significant amount of land has been used for industrial and urban construction. However, the environmental conscientiousness of land operators and related governmental subdivisions cannot keep up with economic development due to the lack of corresponding environmental regulations. The soil and groundwater are polluted during the construction or operation period, leading to contamination. This causes negative effects, such as destruction of the environmental system, increasing health risks, and obstacles to economic development (Hong et al., 2014; Li et al., 2015; Luo, 2009; Qu et al., 2016). Owing to the seriousness of soil pollution, the Chinese government has issued relevant laws and regulations on soil pollution prediction, early warning, emergency management, and construction land information management (Ministry of Ecology and Environmental PRC, 2016; National People's Congress of PRC, 2018), which has forced related enterprises and governmental sectors to make efforts to enhance soil environmental protection and plan environmental information management.

Soil pollution of construction land is a typical data-intensive issue due to various types of pollutants, including heavy metals, such as mercury, cadmium, lead, arsenic, and chromium (Sodango et al., 2018); non-metals such as arsenic and selenium; organic pesticides; oils; phenols; benzodiazepines; and detergents and their derivatives from plant's tailings, waste stone, fly ash, and solid waste (Zhao et al., 2015). These pollutants are released from many sources and related factors, including industrial sewage, domestic wastewater, acid rain, exhaust emissions, accumulations, seawater intrusion, and groundwater transfer.

* Corresponding author. School of Earth Science and Engineering, Sun Yat-sen University, Zhuhai 519082, China.
E-mail address: zhouyz@mail.sysu.edu.cn (Y. Zhou).

https://doi.org/10.1016/j.jenvman.2022.115685
Received 12 March 2022; Received in revised form 24 June 2022; Accepted 3 July 2022
Available online 8 July 2022
0301-4797/© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Secondary diffusion under natural conditions forms a wider range of contamination while the soil in the accumulation site is directly contaminated (Chen et al., 2015). Furthermore, the amount of data collected in the process of soil pollution monitoring, prevention, and early warning is huge, with features of multi-source heterogeneity, fuzzy relationships, and numerous associations (McBratney et al., 2014). In the domain of polluted soil improvement, Okpara et al. (2020) indicated that resources, data availability and uncertainty generated the complexity of decision making in ameliorating soil threats. Therefore, scientific analysis and management of urban soil pollution is an extremely complex task. Zhou et al. (2021) have pointed out that it is necessary to organize the data, relationships, rules, logic, knowledge, and models of urban soil pollution processes into a multi-scale, multi--temporal database, to construct models of urban soil pollution targets with intelligent monitoring and control. However, some shortcomings, such as failure to meet the requirements of monitoring, control, and early warning of urban soil pollution in traditional physical modeling and statistical methods and lack of intelligent systems based on big data analysis have been reported, thus affecting the further application of the huge amount of information regarding pollution.

The Knowledge Graph, originally proposed in 2013, utilizes data science and artificial intelligence, together with big data and deep learning (Ji et al., 2021). Knowledge graph (KG) technology can obtain structured knowledge from massive data to attain unified expression and efficient storage, and it is an effective means for solving the problem of management of massive data in the soil pollution process. Some studies have implemented KG technology to solve environmental problems. Du et al. (2016) constructed an ontology database for describing soil properties such as soil strength, and soil processes such as soil compaction, to address challenges in combining knowledge and expertise in multiple areas while assessing the value of soil environment in the UK. Sermet and Demir (2018) designed an intelligent system, aimed at improving public preparedness for natural disasters, by building a knowledge engine containing voice recognition and natural language processing based on a generalized ontology that extracts data from environmental observations, forecast models, and disaster knowledge databases. Fang et al. (2020) developed a KG with computer vision algorithms to identify hazards on construction sites.

Few studies have been conducted on urban soil pollution based on KGs. Therefore, considering the significant efficiency of KGs in information storage, query, discovery, and reasoning, this study attempts to construct a KG for contaminated construction sites with functions of pollution information queries, relationship reasoning, and knowledge updates to explore the application of KG technology in soil pollution prevention, control, and risk assessment in urban areas.

In this study, we transform the massive heterogeneous data related to soil pollution in urban plots into graph data composed of triples to attain unity of storage and efficient display, mine the hidden information in mass data using graph data algorithm, and explore the feasibility and effectiveness of KG in soil environmental management.

2. Material and methods

2.1. Knowledge graph framework

A KG is defined as an interconnected dataset enriched with semantics that can reason the underlying data for complex decision-making (Natarajan, 2020). A KG transforms the intricate document data into numbers of simple triplets consisting of "entity–relationship–entity" to achieve rapid response and reasoning for a large amount of information.

The basic structure of a KG consists of three main components, nodes, edges, and labels (Fig. 1). Any object, such as a place or person can be a node, presenting an entity or its property in KG, edges represent the relationships between nodes, and labels are used to identify different nodes or edges (IBM Cloud Education, 2021).

The typical process of KG construction is as follows. 1) Exact information from various structured (panel data, row table), semi-structured (rational database, NoSQL), or unstructured data (documental report, text, images, videos), which can be divided into entity, property, and relationship. 2) Present the extracted knowledge as interconnected triplets (head entity, relationship, tail entity) in the format of RDF, JSON, or attributed graph, 3) Integrate the constructed KG into a computational framework of applications in the real world, such as natural language processing, question and answer systems, multi-hop reasoning, and recommender systems (Ji et al., 2021). The process of KG construction is shown in Fig. 2.

2.2. Study area

This study was conducted in over 2000 sites from 14 municipal cities in southern China, the most developed region in the country. And the information collected from the area included investigating reports, official government documents, yearbooks, and related websites.

2.3. Steps in knowledge graph construction

This study includes data mining and extraction, ontology building, and KG construction, which are specifically described below.

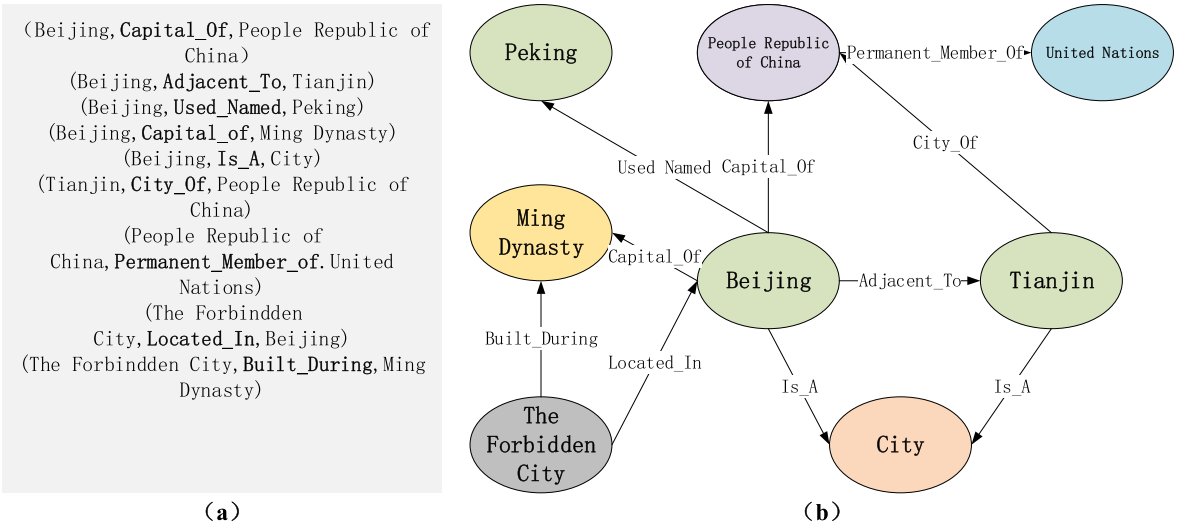


Fig. 1. Example of triplets and knowledge graph (KG). (a) Factual triplets in KG. (b) Entities and relations in KG.

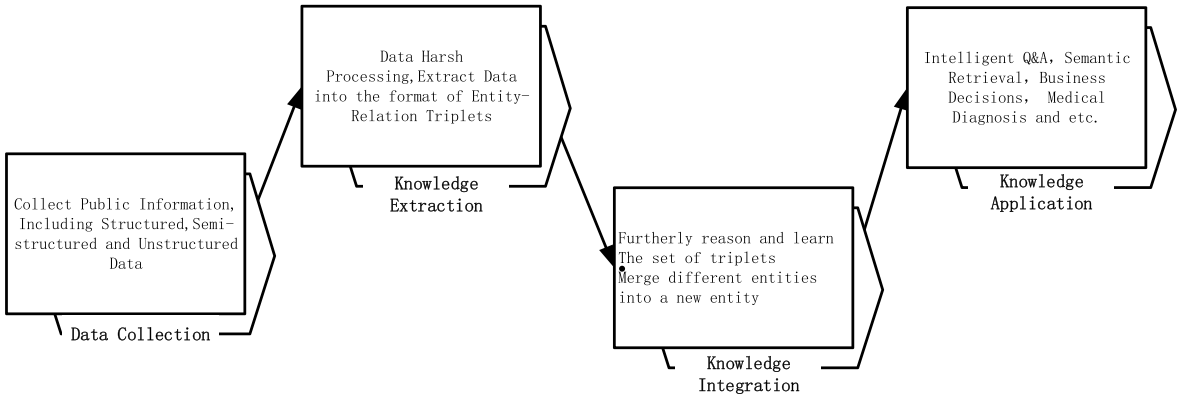


Fig. 2. Process of knowledge graph construction.

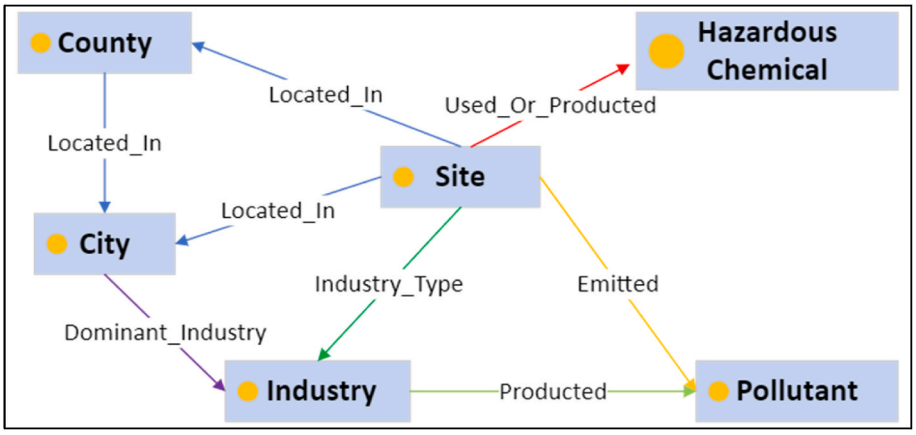


Fig. 3. Ontology of contaminated sites.

2.3.1. *Knowledge extraction*

Knowledge extraction is essential for KG construction. This study collected information from various channels, including site investigation reports, regional yearbooks, government work reports, and websites. The research material we selected and their source are listed in Table 1. Then all the collected data was converted into the triplet format with sites, chemicals, pollutants, and cities as entities and defined connections as different relationships.

2.3.2. *Ontology model building*

Ontologies represent the backbone of formal semantics of KG. They can be seen as data schemes of the graph, and building ontology is the essential of designing a computable knowledge model to semantically represent, organize, and utilize the diversified resource (An et al., 2020). It has been proven that ontologies can serve as a conceptual and technological representation of large information models to enable semantic interoperability (Hildebrandt et al., 2020; Ismail et al., 2012) because ontologies can be stored in a machine-interpretable format and they can work as tools for querying, inference, and visualization. Ontology building is a process that contains structural and logical complexities while its methods can be automatic, semi-automatic, or manual. Automatic and semi-automatic methods heavily rely on deep learning

Table 1
List of data sources for KG construction of contaminated sites.

Data Source	Provider	Description	Size of Data	Method of Collection
Basic information on the studied sites	Institute of Environmental Management and Academy of Environmental Science	Position coordinates, enterprises on sites, pollutant emission, and planning layout.	Over 2000 sites with 86 properties of each site	Extracting information from environmental survey reports of sites, converting into statistical sheets, and connecting the sheets to relational databases
Socio-economic data of the municipal cities and counties	Yearbook published by the provincial and municipal bureaus of statistics	Covering key statistical indicators at local city and county levels, including economy, population, lands employment, energy, resource, and environment.	Over 20 indicators for 14 cities and over 70 counties	Extracting information in the yearbook and statistical bulletin and then converting it into panel data
Information on chemicals consumed or produced at the studied sites	Professional websites and online encyclopedia	Toxicity and physicochemical characteristics of chemicals	Panel data of more than 80,000 chemicals with 30 physical and chemical properties	Acquiring data from related websites and databases, such as *iChemistry* and *Baidu Encyclopedia,* using Python
Spatio-temporal information on the studied sites	Vector maps and remote sensing data	Distance between the sites and sensitive receptors and hydrological information around the sites.	Over related 20 indicators for each studied site	Extracting information from maps by means of spatial measurement using the MapGIS software

algorithms to exact entities and design rules from context, which need a large amount of label data for support (Mannes and Golbeck, 2007; Rezgui, 2007). Currently, the most common method of ontology construction, especially in the professional domain, is manual editing, especially in the research area that lacks sufficient label data. This study defined the label, type, and relationship of entities in urban soil environment using expert experience and literature review (Schwaller et al., 2021; Zhao et al., 2015). Meanwhile, several existing ontologies of general knowledge or professional domain, such as the ontology of soil property, were used for reference; the process was constructed by Du et al. (2016) and Microsoft Concept Graph (Zhang and Yan, 2017).

2.3.3. KG construction

The final step of KG construction was converting the collected information from various sources into triplets, (Zhou et al., 2021), using the Python script while the entities were attached with properties. Since all the material was transformed into structured data, the process of construction was a simple task. The triplets were stored in the format of graph database, and they were visualized using Neo4j, whose query language is Cypher. For more information on the implementation steps of KG construction, refer to this code in github (https://github.com/Feng-David/Tranform-the-information-of-sites-into-neo4j.git) (see Fig. 3).

3. Results and discussion

3.1. Overview of the KG of contaminated sites

The study was conducted in approximately 2500 sites in14 municipal cities located in 75 counties. More than 500 types of industries were distributed in the research sites, and the sites had produced or used 20,644 chemicals, categorized into 1012 pollutants. Each site was assigned a score of environmental risk by experts according to national standards (Ministry of Ecology and Environment, 2018). Over 20,000 entities and 10,000 relations were embedded in the KG model in this study, as shown in Table 2. The visualized graph of the model is presented in Fig. 4.

3.2. Application of knowledge graphs

A KG can be seen as a super knowledge base containing massive data with semantics for humans and computability for machines, which can be relied on to search content. Further, the search of existing KGs can be extended to obtain new knowledge and conclusion (Chen et al., 2020). This study attempted to explore the two typical functions of KG model and their application in contaminated sites.

3.2.1. Information inquiry

Because searching the required information from massive data is the strength of KG technology (Ji et al., 2021), this study used constructed KGs to search the information of contaminated sites. A KG was constructed using Neo4j software for graph databases, whose search language is Cypher. This study used the *MATCH* clause (Neo4j, 2020) in Cypher to search basic information of a certain site, and the result was shown in Fig. 5 A. The related chemicals, pollutants, and its score of environmental risk were all present as connected triplets. Further query can be achieved by introducing more complex rules and syntax. This study used the arsenic pollutant to inquire the sites that emitted arsenic and the industry type and assessment of environmental risk of sites (Fig. 5b).

3.2.2. Knowledge reasoning

The definition of knowledge reasoning based on KG is inferring unknown facts or relations based on existing facts or relations in the graph and generally focusing on the characteristic information of entities, relations, and graph structure. Specifically, KG reasoning can help deduce new facts, new relations, new axioms, and new rules (Shi and Weninger, 2018). This study used the knowledge reasoning function to explore the hidden information of constructed KG of contaminated sites. The cases of application in this study were discovering the characteristic pollutants of certain industries and finding out the regional distribution of pollutants. A typical example was discovering the characteristic pollutants of Metal Surface Manufacturing industry using information in databases. The process was as follows (Fig. 6). 1) Transform the database into KG, stored as nodes and edges. 2) Display all sites linking to the node of the industry, which is called Metal Surface Manufacturing and all pollutants linking to the sites. 3) Count the number of links between the sites of each pollutant. 4) Rank pollutants by the number of links. Finally, the characteristic pollutants of the Metal Surface Manufacturing industry were determined to be benzopyrene, arsenic, and ammonia.

Another case of application of knowledge reasoning was exploring the regional distribution of pollutants. The arsenic pollutant can be a case in this study. The process was similar to that mentioned above in the case of searching the characteristic pollutant of the industry. 1) Display the nodes of sites linking to the arsenic pollutant, nodes of cities linking to the related sites, and their links. 2) Count the links of the sites of each city. 3) Rank cities by the number of links and presenting them as a histogram. The result is shown in Fig. 7.

3.3. Significant findings

This study transformed numerous multi-source heterogeneous data into graph data that consist of nodes and edges, so that the hidden information can be detected and the complex problem can be transferred into mathematical problem and solved by high-performant computing devices.

This article summarized the organizational and technological process and main value of the graph database in detecting environmental risk. The result demonstrated the construction of KG model for urban soil environment based on instructed data. The huge amount of instructed data, including survey reports and remote sensing images, were the significant material for sematic network. Furthermore, compared to the traditional relational database, the KG model in the form of graph for data storage has advantages such as simple

Table 2
Overview of the entities and relations of the knowledge graph of contaminated sites.

Type of Entity	Quantity of Entity	Type of Relation	Name of Relation	Quantity of Entity
Site	2492	Site—>City	Located In	2492
City	21	Site—>County	Located In	2492
County	75	Site—>Industry	Industry Type Is	2453
Industry	519	Site—>Chemical	Produced or Used	29,795
Chemical	20,644	Site—>Pollutant	Characteristic Pollutant Is	8333
Pollutant	1012	Site—>Pollutant	Excessively Emitted	62
Score of Environmental Risk	1672	Site—>Score of Environmental Risk	Score Is	2492
		County—>City	Located In	75
		County—>Industry	Leading Industry Is	87
		Chemical—>Pollutant	Subclass of	25,031

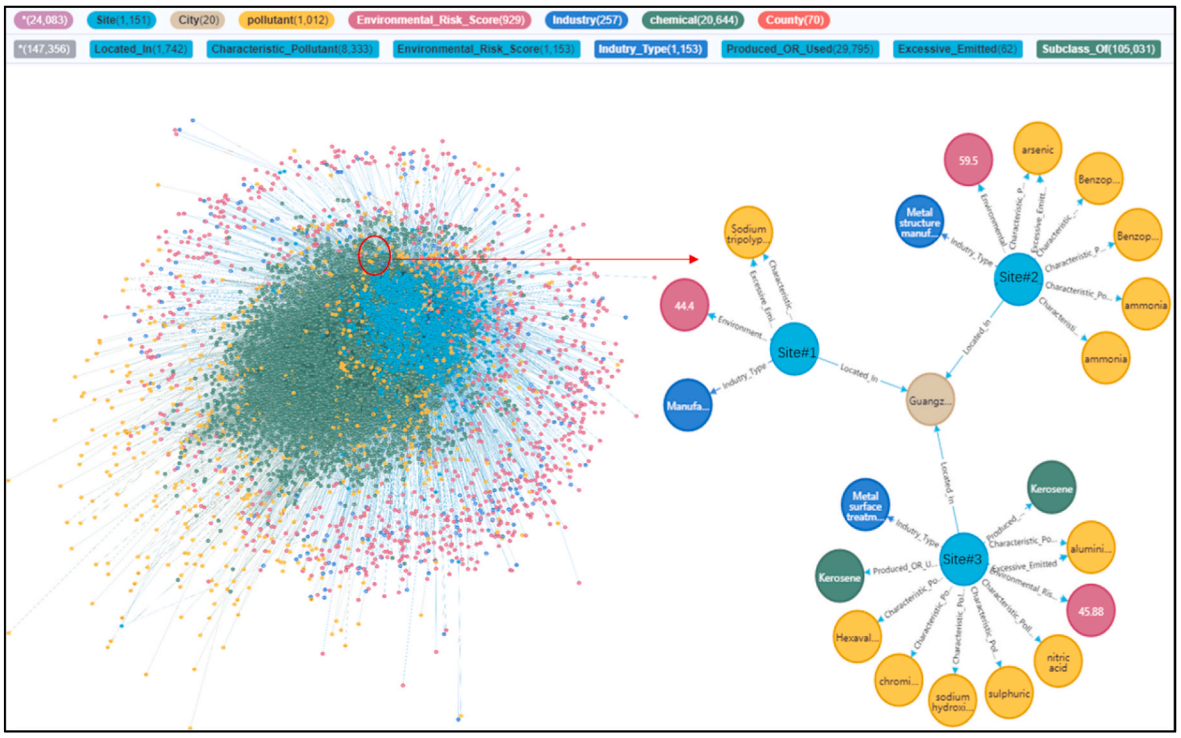


Fig. 4. General view of the knowledge graph of contaminated sites.

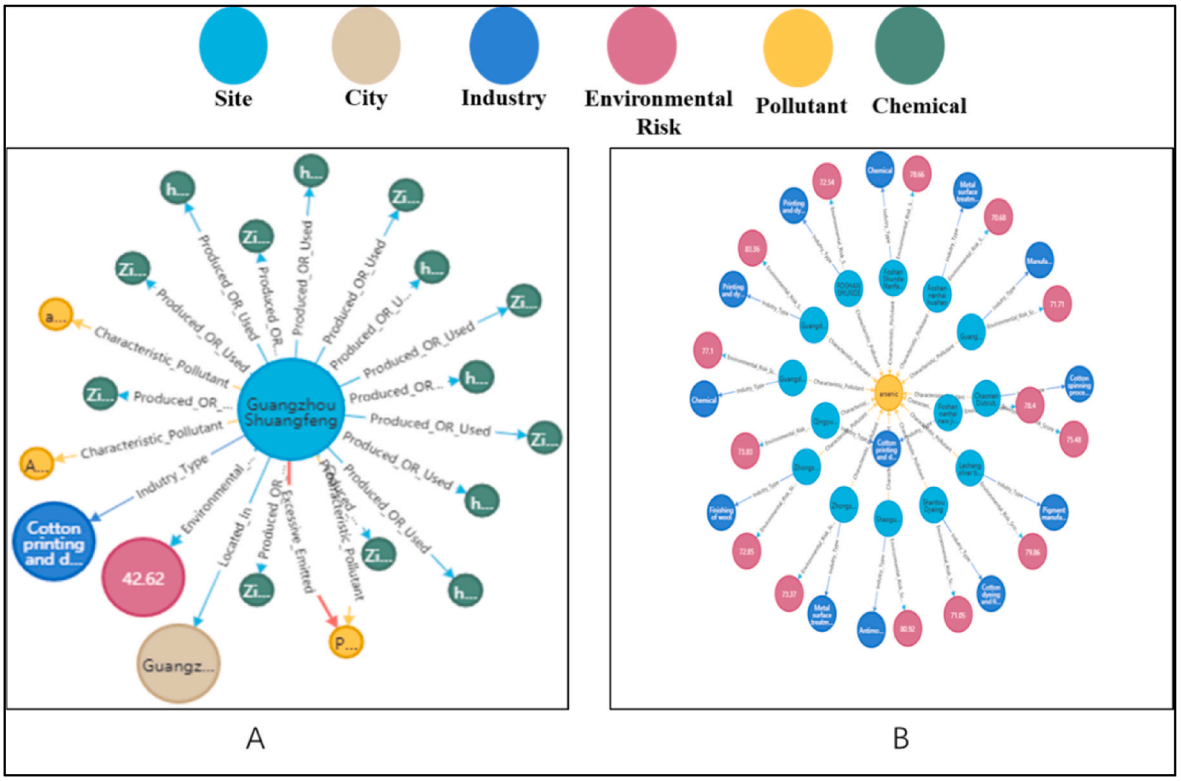
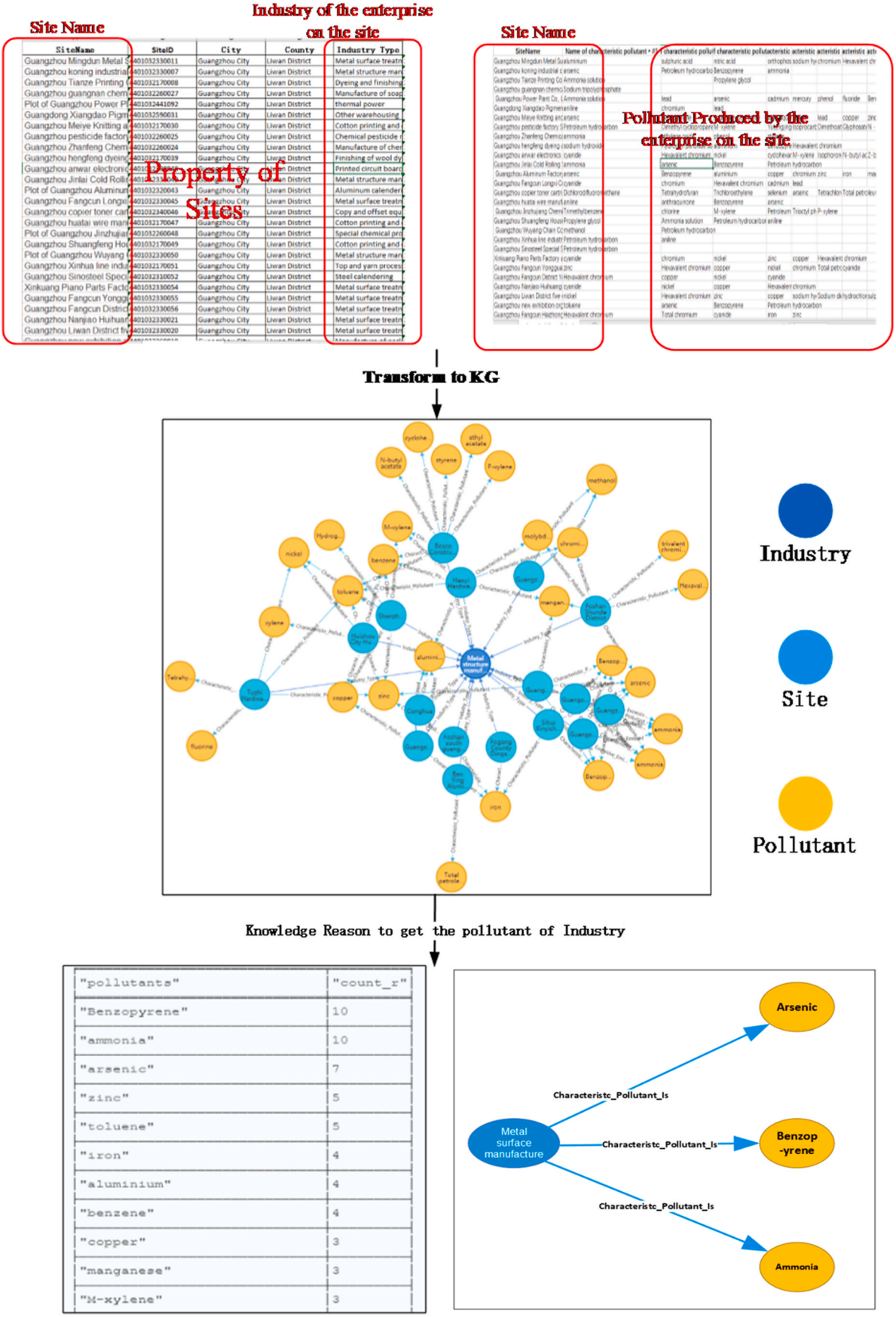


Fig. 5. Results of searching the knowledge graph. (a) Related information on arsenic. (b) Site information graph.

representation of complex data; flexible change in the relationships among entities; determination of the relationship between two nodes on one's own will; implementation of graph algorithms such as community discovery algorithm, mediation center algorithm, and some algorithms involved in the routing of computer networks to assist information query (Ghrab et al., 2016). Therefore, the constructed KG of contaminated sites can serve as an efficient framework of numerous related information to enhance the capability of management of soil environment.

For further implementation of KG, some algorithms of knowledge retrieval, knowledge discovery, and knowledge intervention work as



"pollutants"	"count_r"
"Benzopyrene"	10
"ammonia"	10
"arsenic"	7
"zinc"	5
"toluene"	5
"iron"	4
"aluminium"	4
"benzene"	4
"copper"	3
"manganese"	3
"M-xylene"	3

Fig. 6. Characteristic pollutants of the industry using knowledge reasoning—metal surface manufacturing as case.

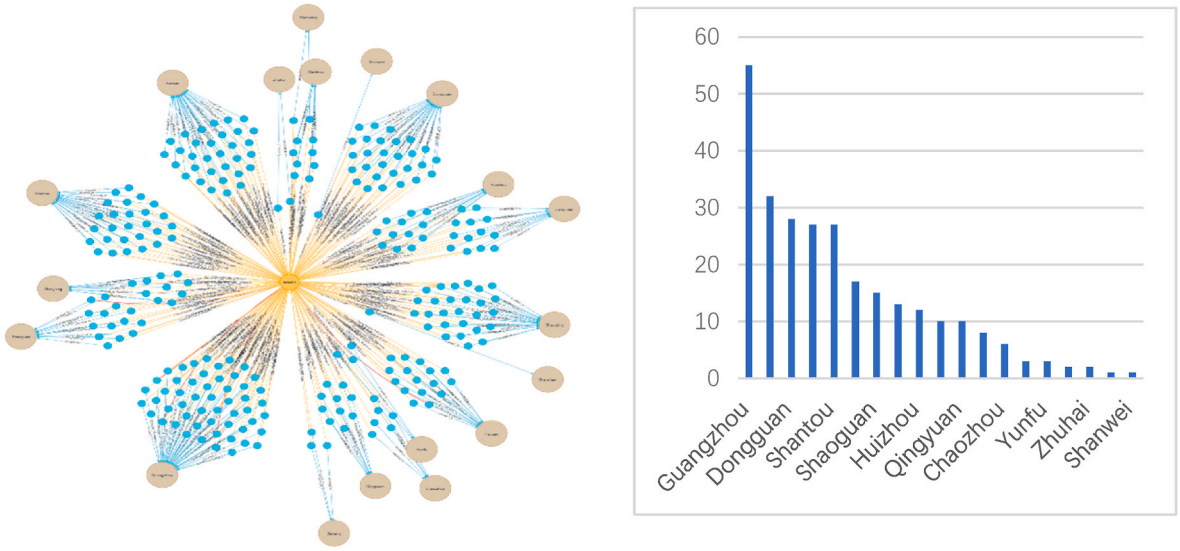


Fig. 7. Distribution of the pollutant via knowledge reasoning—arsenic as the case.

efficient tools to reveal hidden information. In this study, a simple knowledge reason algorithm was used to find featured pollutants of certain industries and regions. The cases proved that KG technologies can quickly and precisely identify the level of environmental risk of the studied sites. However, KG algorithms can help find out related factors that lead to soil pollution such as the industry type, factory layout, background contents of heavy metals in different soils, and knowledge discovery through KG technologies in pollution detection and early warning.

From the perspective of contaminated site management, KG is the preferred method to analyze and visualize the complete information of sites, which can assist in discovering and tracing the hidden risks in health and socioeconomic issues. The study results demonstrated that the information search in KG was compatible with multi-source heterogeneous environmental and chemical big data; thus, it could accurately and rapidly disclose the concealed environmental problems in contaminated sites and their surrounding areas, which is of great significance in optimizing decision-making in soil environmental management.

Finally, the visualization of KG in contaminated sites, presented as a large social network, consisted of a number of triplets while the relationship among entities was clearly shown, which enhanced the readability and computability for machine and intuitive interpretability for readers of different kinds of data related to the soil environment.

3.4. Limitations and prospective improvements

The main limitation was that this study was based on small samples and insufficient information of contaminated sites. The number of samples selected in this study only accounted for a small part of the number of construction sites. Therefore, if the number of sites increased to a large scale, it is doubtful whether the current KG model can ensure certain stability, accessibility, and accuracy in the process of information storage, retrieval, and inference (Weis and Jacobson, 2021).

On the other hand, A large portion of the semantic material to construct the KG model was derived from structured data such that the process of KG construction was relatively easier than that based on multimode data (Bloem et al., 2021; Deng et al., 2021). And in the process of knowledge retrieval and calculation, it is necessary to replace Cypher language with natural language to reduce the difficulty of popularization and application of KG technologies in the management of contaminated sites. The optimistic prediction is that a few large models or pre-training models based on deep learning have been developed and implemented in some knowledge domains (Liu and Sun, 2021), which can automatically generate multimodal KGs from large-scale data depending on the requirement of downstream tasks. Currently, several natural language process training models for the Chinese corpus have been published and applied to some knowledge domains, including classical poetry (Cui et al., 2021), fairy tales (Lai et al., 2021), and bio-medicine (Zhang et al., 2020). These cases indicate the feasibility of automatic entity extraction and KG construction.

In addition, data security issue must be considered in the KG model. Under China's current legal system, soil pollution status is sensitive information. It is necessary to avoid leakage of relevant information of soil pollution to the public. Therefore, all information relating to site contamination must be stored in a closed system. This pattern limits the model's ability to handle large-scale data and data sharing between different departments.

Furthermore, the process of KG construction in this study, including data collection, ontology building, and entity extraction, was not sufficiently intelligent. For example, basic information of the studied sites was exacted from detailed survey reports of soil environment of the sites, which were firstly manually transformed into the format of panel data and then stored as a graph dataset by staff. This limitation is not only in environmental science issue, Nicholson and Greene (2020) concluded that utilization of automated system was scarce during biomedical knowledge graph construction. Thus, the method of KG construction was typically inefficient and required large numbers of human intervention. Therefore, how to achieve automatic and intelligent information extraction, especially in the issue of soil environment should be the focus of future studies.

4. Conclusion

This work aims to solve the problem of information stored, management and utilization of contaminated site which was faced by the agents of environmental protection. Based on the feature of KG and graph data algorithms, this study selected 2000 sites in the most developed regions of China as cases to explore the feasibility and accessibility of KG construction and application in urban contaminated sites and soil environmental management. The results indicated the great potential of KG technology for management of contaminated sites and prediction of soil pollution. And the graph algorithm can help reveal complete information within and around certain sites and contract the hidden risks, which is beneficial for government agencies to implement precise environmental emergent event management measures and

provide rapid, evidence-based decision support to enhance the responsiveness of urban environmental emergency response system.

However, in order to achieve the intelligent information query and convenient user interaction, further studies are need. The shortage, including how to construct KG model from multimode data, how to realize the interaction between users and KG system through natural language, and how to ensure the rapidity and stability when processing massive data, still need further exploration.

Credit author statement

Feng Han: Conceptualization, Methodology, Formal analysis, Writing – original draft. Yirong Deng: Formal analysis, Investigation. Qiyuan Liu: Writing – review & editing, Methodology. Yongzhang Zhou: Conceptualization, Methodology, Review. Jun Wang: Review, Methodology, Data curation. Yongjian Huang: Writing – review & editing. Qianlong Zhang: Writing – review & editing. Jing Bian: Methodology, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Zhou Yongzhang reports financial support was provided by National Natural Science Foundation of China.

Acknowledgements

Funding: This work was supported by the Environmental Academy of Guangdong, Sun Yat-Sen University through the State Key Program funded by the National Natural Science Foundation of China [grant number U1911202) and Guangdong Provincial Key Research Program funded by Guangdong Provincial Department of Science and Technology [grant number 2020B111137001].

References

An, Y., Qin, F., Sun, D., Wu, H., 2020. A multi-facets ontology matching approach for generating PLC domain knowledge graphs. IFAC-PapersOnLine 53, 10929–10934. https://doi.org/10.1016/j.ifacol.2020.12.2834.

Bloem, P., Wilcke, X., van Berkel, L., de Boer, V., 2021. Kgbench: A Collection of Knowledge Graph Datasets for Evaluating Relational and Multimodal Machine Learning. Springer International Publishing, Cham, pp. 614–630.

Chen, H., Teng, Y., Lu, S., Wang, Y., Wang, J., 2015. Contamination features and health risk of soil heavy metals in China. Sci. Total Environ. 512, 143–153. https://doi.org/10.1016/J.SCITOTENV.2015.01.025.

Chen, X., Jia, S., Xiang, Y., 2020. A review: knowledge reasoning over knowledge graph. Expert Syst. Appl. 141, 112948 https://doi.org/10.1016/j.eswa.2019.112948.

Cui, Y., Che, W., Liu, T., Qin, B., Yang, Z., Wang, S., Hu, G., 2021. Pre-training with whole word masking for Chinese BERT. IEEE/ACM Transact. Audio Speech Lang. Process. 29, 3504–3514.

Deng, C., Jia, Y., Xu, H., Zhang, C., Tang, J., Fu, L., Zhang, W., Zhang, H., Wang, X., Zhou, C., 2021. GAKG: a multimodal geoscience academic knowledge graph. In: Proceedings of the 30th ACM International Conference on Information & Knowledge Management. Association for Computing Machinery, pp. 4445–4454.

Du, H., Dimitrova, V., Magee, D., Stirling, R., Curioni, G., Reeves, H., Clarke, B., Cohn, A., 2016. An Ontology of Soil Properties and Processes. Springer International Publishing, Cham, pp. 30–37.

Fang, W., Ma, L., Love, P.E.D., Luo, H., Ding, L., Zhou, A., 2020. Knowledge graph for identifying hazards on construction sites: integrating computer vision with ontology. Autom. ConStruct. 119 https://doi.org/10.1016/j.autcon.2020.103310.

Ghrab, A., Romero, O., Skhiri, S., Vaisman, A., Zimányi, E., 2016. GRAD: on graph database modeling. arXiv e-prints arXiv:1602.00503.

Hildebrandt, C., Köcher, A., Küstner, C., López-Enríquez, C.M., Müller, A.W., Caesar, B., Gundlach, C.S., Fay, A., 2020. Ontology building for cyber–physical systems: application in the manufacturing domain. IEEE Trans. Autom. Sci. Eng. 17, 1266–1282. https://doi.org/10.1109/TASE.2020.2991777.

Hong, Y., Huang, X., Thompson, J.R., Flower, R.J., 2014. China's soil pollution: urban brownfields. Science 344, 691–692. https://doi.org/10.1126/science.344.6185.691-b.

IBM Cloud Education, 2021. What is a knowledge graph? https://www.ibm.com/cloud/learn/knowledge-graph. (Accessed 12 April 2021).

Ismail, H., Hegazy, A., Badr, A., Gawich, M., 2012. A methodology for ontology building. Int. J. Comput. Appl. 56, 39–45. https://doi.org/10.5120/8867-2834.

Ji, S., Pan, S., Cambria, E., Marttinen, P., Yu, P.S., 2021. A survey on knowledge graphs: representation, acquisition, and applications. IEEE Transact. Neural Networks Learn. Syst. 1–21. https://doi.org/10.1109/TNNLS.2021.3070843.

Lai, Y., Liu, Y., Feng, Y., Huang, S., Zhao, D., 2021. Lattice-BERT: leveraging multi-granularity representations in chinese pre-trained language models. ArXiv abs/2104.07204.

Li, X., Jiao, W., Rongbo, X., Chen, W., Chang, A., 2015. Soil pollution and site remediation policies in China: a review. Environ. Rev. 23, 150318143344008 https://doi.org/10.1139/er-2014-0073.

Liu, W., Sun, S., 2021. Awesome Pretrained Chinese NLP Models. Github, Hangzhou.

Luo, Y.M., 2009. Trends in Soil Environmental Pollution and the Prevention-Controlling-Remediation Strategies in China. Environmental Pollution & Control.

Mannes, A., Golbeck, J., 2007. Ontology building: a terrorism specialist's perspective. IEEE Aerospace Conference Proceedings. https://doi.org/10.1109/AERO.2007.352794.

McBratney, A., Field, D.J., Koch, A., 2014. The dimensions of soil security. Geoderma 213, 203–213. https://doi.org/10.1016/J.GEODERMA.2013.08.013.

Ministry of Ecology and Environment, PRC., 2018. National environmental protection standards of the people's Republic of China. In: Technical Guideline for Vertification of Risk Control and Soil Remedation of Contaminated Site. Ministry of Ecology and Environment, PRC;, Beijing, p. 10 (in Chinese).

Ministry of Ecology and Environmental the People's Republic of China, 2016. Action plan targets on soil pollution, pp. 1–2. http://www.gov.cn/zhengce/content/2016-05/31/content_5078377.htm.

Natarajan, M., 2020. From graph to knowledge graph: a short journey to unlimited insights. In: Neo4j (Ed.), Neo4j Resource Library. Neo4j (San Mateo, California).

National People's Congress of the People's Republic of China, 2018. Law of the People's Republic of China on Prevention and Control of Soil Contamination, pp. 2–10. http://www.npc.gov.cn/zgrdw/npc/lfzt/rlyw/node_32834.htm. (Accessed 1 January 2019).

Neo4j, I., 2020. The Neo4j Cypher Manual v4.3, 3 ed. Neo4j, Inc, p. 4. Neo4j, Inc.

Nicholson, D.N., Greene, C.S., 2020. Constructing knowledge graphs and their biomedical applications. Comput. Struct. Biotechnol. J. 18, 1414–1428. https://doi.org/10.1016/j.csbj.2020.05.017.

Okpara, U.T., Fleskens, L., Stringer, L.C., Hessel, R., Bachmann, F., Daliakopoulos, I., Berglund, K., Blanco Velazquez, F.J., Ferro, N.D., Keizer, J., Kohnova, S., Lemann, T., Quinn, C., Schwilch, G., Siebielec, G., Skaalsveen, K., Tibbett, M., Zoumides, C., 2020. Helping stakeholders select and apply appraisal tools to mitigate soil threats: researchers' experiences from across Europe. J. Environ. Manag. 257, 110005 https://doi.org/10.1016/j.jenvman.2019.110005.

Qu, C., Shi, W., Guo, J., Fang, B., Wang, S., Giesy, J.P., Holm, P.E., 2016. China's soil pollution control: choices and challenges. Environ. Sci. Technol. 50, 13181–13183. https://doi.org/10.1021/acs.est.6b05068.

Rezgui, Y., 2007. Text-based domain ontology building using Tf-Idf and metric clusters techniques. Knowl. Eng. Rev. 22, 379–403. https://doi.org/10.1017/S0269888907001130.

Schwaller, P., Hoover, B., Reymond, J.L., Strobelt, H., Laino, T., 2021. Extraction of organic chemistry grammar from unsupervised learning of chemical reactions. Sci. Adv. 7 https://doi.org/10.1126/sciadv.abe4166.

Sermet, Y., Demir, I., 2018. An intelligent system on knowledge generation and communication about flooding. Environ. Model. Software 108, 51–60. https://doi.org/10.1021/acs.est.6b0506810.1016/j.envsoft.2018.06.003.

Shi, B., Weninger, T., 2018. Open-World Knowledge Graph Completion. AAAI, p. 8.

Sodango, T.H., Li, X., Sha, J., Bao, Z., 2018. Review of the spatial distribution, source and extent of heavy metal pollution of soil in China: impacts and mitigation approaches. J. Health Pollut. 8, 53–70. https://doi.org/10.5696/2156-9614-8.17.53.

Weis, J.W., Jacobson, J.M., 2021. Learning on knowledge graph dynamics provides an early warning of impactful research. Nat. Biotechnol. 39, 1300–1307. https://doi.org/10.1038/s41587-021-00907-6.

Zhang, D., Yan, J., 2017. Microsoft Concept Graph. Data Mining and Enterprise Intelligence Group. MSRA. https://concept.research.microsoft.com/Home/Demo.

Zhang, N., Jia, Q., Yin, K., Dong, L., Gao, F., Hua, N., 2020. Conceptualized Representation Learning for Chinese Biomedical Text Mining. arXiv e-prints arXiv:2008.10813. https://doi.org/arXiv:2008.10813.

Zhao, F.-J., Ma, Y., Zhu, Y.-G., Tang, Z., McGrath, S.P., 2015. Soil contamination in China: current status and mitigation strategies. Environ. Sci. Technol. 49, 750–759. https://doi.org/10.1021/ES5047099.

Zhou, Y., Zhang, Q., Huang, Y., Yang, W., Xiao, F., Ji, J., Han, F., Tang, L., Ouyang, C., Shen, W., 2021. Constructing knowledge graph for the porphyry copper deposit in the Qingzhou-Hangzhou Bay area: insight into knowledge graph based mineral resource prediction and evaluation. Earth Sci Front. 28, 67–75. https://doi.org/10.13745/j.esf.sf.2021.1.2.

View publication stats