ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume X-3/W2-2022
Urban Geoinformatics 2022, 1–4 November 2022, Beijing, China

RESEARCH ON THE CONSTRUCTION OF GEOGRAPHIC KNOWLEDGE GRAPH INTEGRATING NATURAL DISASTER INFORMATION

Binyu Zhang [1], Chuan Yin [1,*], Kexin Liu [1], Xi Zhai [2], Yuqing Sun [3], Mingyi Du [1,*]

[1]School of Geomatics and Urban Spatial Informatics, Beijing University of Civil Engineering and Architecture, Beijing 102616, China-2108160220006@stu.bucea.edu.cn (B. Zhang), yinchuan@bucea.edu.cn (C. Yin), kxliukexin@163.com(K. Liu),dumingyi@bucea.edu.cn(M. Du)
[2]Information Service Department, National Geomatics Centre of China, Beijing 100830, China-zhaixi@ngcc.cn
[3]State Key Laboratory Incubation Base of Urban Environmental Processes and Digital Simulation, Capital Normal University, Beijing 100048, China-syuq2000@163.com

KEY WORDS: Natural Disasters, Geographic Knowledge Graph, Knowledge Service, Knowledge Engineering, Natural Language Processing, Linked Data.

ABSTRACT:

Natural disasters have a significant impact on the environment and economies of all countries around the world, and a large amount of multi-source heterogeneous geographic information data is generated every day. However, due to a lack of knowledge transformation capabilities, these nations continue to struggle with the issue of "a large amount of data and little knowledge". Therefore, it is of great significance how to extract geographic knowledge related to disasters from the vast data and construct a geographic knowledge graph integrating disaster information. Based on the theory related to knowledge extraction, this paper proposes a method to construct a natural disaster knowledge graph integrating geographic information. The core of this knowledge graph is to construct the association relationship between natural disaster concepts, research areas, and spatial data. The vocabulary and relationships associated with disaster concepts are primarily transformed by an existing word list of geographic narratives, which then provide rich semantic relationships of domain concepts for the entire knowledge graph. The research areas and spatial data types are mainly obtained through knowledge entity extraction and disambiguation methods. This disaster knowledge graph can support applications well such as natural disaster visualization and analysis, data recommendation systems, and intelligent Q&A systems, which can further improve the intelligence of natural disaster knowledge services and is expected to promote the sharing and reuse of domain knowledge graphs to a certain extent.

1. INTRODUCTION

After years of work, most countries around the world have built a basic geographic information database system with multiple scales, rich content, and timely updates, which can provide data and technical support for geological disasters, floods, forest fires, and other emergency emergencies. In recent times, with the quiet rise of basic geospatial knowledge services, we are required to not only produce data but also provide knowledge services to disaster emergency managers, such as spatialized knowledge graph services and spatial data recommendation services, to support their planning, management, and decision-making studies. However, we are currently unable to meet the service demand of users for basic geographic knowledge. The reason is that the above knowledge mainly exists implicitly in the unstructured form in various types of books, journal papers, dissertations, science and technology reports, patent descriptions, and other documentary resources in different literature, and the knowledge contained in these literature carriers can neither be used by automated systems nor managed by people in a very convenient way. This means that knowledge is difficult to be accessed, shared, and reused, resulting in the situation of "massive data, explosive information, and hard to find knowledge", which restricts the full play of the role of basic geographic data and information, and becomes a major problem in the innovation of mapping science and technology(Yu et al. 2020).

The above problem is well-served by knowledge extraction. Knowledge extraction extracts the knowledge points (also called knowledge elements) contained in a document one by one by analyzing and processing the content of the document, marks the attributes of the knowledge, and then stores them in the knowledge base in a certain form. Knowledge extraction is different from data mining and knowledge discovery. It is an effective way of knowledge acquisition and its sublimation and deepening of information extraction (Hua and Zhang,2010). Knowledge extraction efforts usually require a certain amount of domain knowledge base as support, which can assist in implementing rule building, machine learning, etc. Ontology is a knowledge representation structure that has been used more often in the past, but with the dramatic increase in the amount of processed data, the process of ontology construction is becoming more and more difficult, so academics are now considering the use of a completely new form of knowledge representation--knowledge graphs to describe the existence of real-world knowledge. Currently, successful models and experience in this research have emerged in individual fields (Guo and He,2015). The construction of knowledge graphs often requires a great cost, and since current natural language processing methods are not perfect, it is difficult to obtain more accurate knowledge graphs by fully automated construction methods, while fully manual construction methods guarantee accuracy but require huge labor and time costs, and it is almost impossible to construct larger-scale knowledge graphs completely manually. Therefore, how to reconcile accuracy and efficiency, balance automated methods and manual participation, and build the most accurate knowledge graph in the most efficient way is a major challenge to be solved in building knowledge graphs at present (Yang et al.2018).

* Corresponding author

https://doi.org/10.5194/isprs-annals-X-3-W2-2022-79-2022 | © Author(s) 2022. CC BY 4.0 License.

In summary, based on the urgent demand for geographic information knowledge services in the environment of geographic information big data and transformation and upgrading of mapping technology, this project takes domain knowledge graph-driven geographic knowledge extraction as a breakthrough and attempts to propose an accurate and efficient construction method of domain knowledge graph from the perspective of geographic ontology theory, using a large amount of Chinese journal/paper literature as a data source. Research on key technologies such as knowledge entity identification and knowledge point extraction based on knowledge graph ontology model and knowledge reasoning, and explore innovative models of geographic information knowledge services. It can not only provide a new way to acquire geographic information knowledge for Internet big data, but also is expected to promote the sharing and reuse of domain knowledge graphs to a certain extent, and has important application value for the national "four comprehensive" strategic layout, ecological civilization construction and "going out" strategy.

2. RELATED WORK

A knowledge entity is a terminological entity that expresses a key knowledge point in the professional literature (Wen et al.2018). The extraction of knowledge entity types refers to the generalization and extraction of type-indicating words according to the laws of terminological expressions in specialized literature (Wen et al.2018). For example, "algorithm" in "the shortest path algorithm" and "model" in "support vector machine model". As the first step of knowledge point extraction, identifying the types of knowledge entities in a text aims to be able to further extract their semantic roles and other information by identifying key concepts. Since the research area of the literature is also one of the knowledge points extracted in this project, entity identification includes the identification of geographic entities in addition to knowledge entities, and the research progress of these two parts is analyzed separately below.

Geographic entity recognition is to extract elements with spatial location information from web texts, such as administrative divisions, organizations, and gatehouse addresses. For geographic entity recognition, the method based on rule matching is gradually being replaced by supervised machine learning methods due to their low recall and high cost of constructing patterns (Davies,2013; Twaroch et al.2008). In the evaluation of information extraction systems organized at the 2006 CoNLL conference, the F-values (weighted average of algorithm correctness and recall) of the most advanced named entity detection and classification systems for extracting location names and organization names for articles in the news network were 91.15% and 84.67%, respectively. Geographic named entity recognition is considered a solved problem (Marrero et al.2013). In order to meet specific application requirements, identified geographic entities in web texts need to be associated with their spatial locations in the real world. However, the layer-by-layer abstraction of human cognition and the diversity of expressions lead to many ambiguities in geographic entity positioning, and the problems of geographic entity name disambiguation and fuzzy region modeling need to be addressed. The ambiguity of geographic entity names is manifested by homonyms and multiple names for one place. Disambiguation rules are usually written by linguists, but this approach is not effective in disambiguating location names due to the disadvantages of limited coverage of rules, too short descriptive information for dictionaries, and too fine classification of synonym dictionaries (Buscaldi et al.2008). The use of domain tags can establish semantic associations between words and classify entities effectively, but the domain knowledge base is still incomplete and has limited disambiguation capabilities. With the continuous development and improvement of the encyclopedic knowledge base, it has become a rich source of disambiguation knowledge. And the Internet, as a massive corpus without word meaning annotations, provides richly expressed, rapidly updated, and widely covered background knowledge, and using Internet multi-source knowledge to extract the distribution characteristics of entities has become a new trend in the disambiguation of location names (Lieberman and Samet,2011).

In terms of knowledge entity identification, the current research mainly focuses on knowledge entity identification conducted for the information field (Xu et al.2018; Zhai,2017), such as Qiu et al (2012) classified patent design knowledge into target functional knowledge, action principal knowledge, detail feature knowledge, and location feature knowledge according to its application in innovation design. Besides, some scholars in the medical and agricultural fields have also conducted such studies (Li,2018; Li and Zhang,2018). In general, current research on type extraction of such specialized vocabulary and short knowledge entities is rare, and such research in the field of mapping and geographic information is even scarcer. In terms of knowledge entity identification methods, mainly the artificially constructed knowledge base resources are used for category labeling or type labeling propagation, and the commonly used knowledge bases are mainly Wordnet, Wikipedia, Freebase, etc., as well as the domestic "Baidu Encyclopedia" and the "Synonym Word Forest" of Harbin Institute of Technology, and there are many studies based on such methods, and the results achieved are relatively good (Suchanek et al.2008; Ni et al.2010; Dojchinovski and Kliegr,2013). These methods have a relatively high accuracy rate because they borrow an existing knowledge base constructed manually. However, due to the characteristics of the field named knowledge entities with strong specialization, multiple types, and text specification, such entities cannot be identified using a general knowledge base (Wen et al.2018). Therefore, domain knowledge entity recognition needs to introduce more domain knowledge resources and integrate multiple methods.

The type of knowledge extraction in this project belongs to the field of mapping and geographic information, so the knowledge extraction process needs the support of a knowledge graph in the field of geographic information. At present, representative achievements of foreign knowledge graphs in the field of geographic information include GeoNames Ontology, OSM Semantic Network, LinkedGeoData, GeoWordNet, etc. Relatively speaking, the domestic research on geographic information knowledge graph started late and is still at the level of theoretical research. Lu et al (2017) argued that a geographic knowledge graph is the key to expanding traditional geographic information services to geographic knowledge services, and dissected the key scientific problems that need to be solved urgently for the construction of the geographic knowledge graph in terms of semantic understanding of geographic information and spatial semantic computational models. Jiang et al (2018) proposed the construction process of the geographic knowledge graph in conjunction with the rapid conversion and fusion of geographic knowledge in virtual geographic environment systems and explored the application method of the geographic knowledge graph. Based on resources such as subject textbooks and Baidu encyclopedia texts, Yang et al (2018) constructed a geography subject knowledge graph with completed applications and basic education using crowdsourced semi-automatic semantic annotation.

At the present, the difficulty of geographic entity recognition is location name disambiguation, and the use of the multi-source Internet knowledge base for semantic and distributional relationship features of geographic entities has become a new trend of location name disambiguation. Research on knowledge entities is in its initial stage, and the recognition of such entities is difficult, which requires the combination of various recognition methods (such as machine learning, the method based on rule matching, etc.), and the integration of more domain knowledge resources in the recognition process. The automatic construction of general knowledge graphs has made great progress, but the lack of linguistic information such as ontologies in vertical domains has led to slow progress in the construction of domain knowledge graphs, and it is difficult to guarantee the accuracy of the graphs with fully automated construction, while fully manual construction requires huge labor and time costs. Therefore, coordinating accuracy and efficiency, balancing automated methods and human participation, and making good use of existing domain knowledge base resources (such as professional narrative lists and glossaries) to build the most accurate knowledge graph in the most efficient way is a major challenge to be solved in building knowledge graphs at present.

3. METHODS

In this paper, the ontology involved in the geographic knowledge graph mainly includes literature, keywords (natural disaster field), research area, and spatial data, among which literature is directly related to spatial data, research area, and keywords, while spatial data, keywords, and research area are indirectly related through the same literature, and this indirect relationship will increase the credibility with the increase of the number of aggregated literature. For example, if we retrieve articles on the topic of "Mudslide" from the massive literature database and count the provinces and spatial data with high frequency, the semantic relationship between "Mudslide" and counted provinces and spatial data will have high. The semantic relationship of the ontology layer of the knowledge graph is shown in Figure 1.

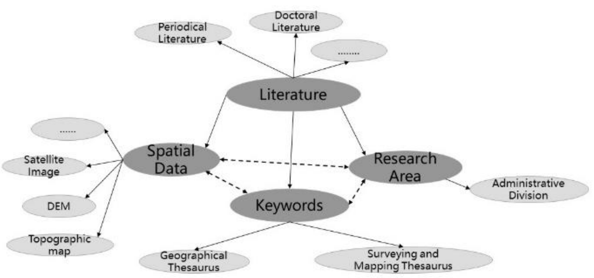

Figure 1. Knowledge graph ontology layer semantic relations.

The keywords of the literature in Figure 1 do not need to be extracted, but the semantic relationships of the keywords (such as contextual relationships) cannot be obtained directly from the literature and can be extracted and converted based on the conceptual relationships in the existing professional thesaurus. The research area and spatial data are mainly obtained through the knowledge entity extraction method based on literature abstracts, and the overall technical route is shown in Figure 2.

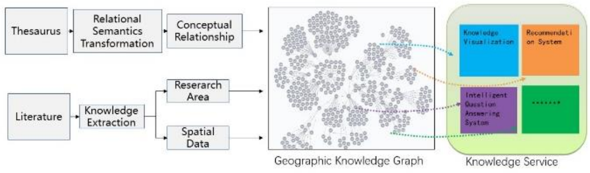

Figure 2. Overall technology road map.

3.1 Disaster concept relationship extraction

The conceptual relationships of the knowledge graph in this research are mainly based on the Earth Science Narrative Word List and the Mapping Science Narrative Word List for relationship extraction. In order to improve the operability and practicality of the conceptual relationships, this paper draws on the experience of converting domestic and foreign knowledge organization systems to SKOS (Simple Knowledge Organization System), based on the core vocabulary and mapping vocabulary of SKOS, which provides a set of standardized, Based on SKOS core vocabulary and mapping vocabulary, SKOS provides a standardized, flexible, simple and scalable description transformation mechanism for knowledge organization systems such as narrative word lists, subject headings, taxonomies and terms, which can be compatible with traditional knowledge organization systems to the maximum extent. Figure 3 shows the flow chart of SKOS transformation technology.

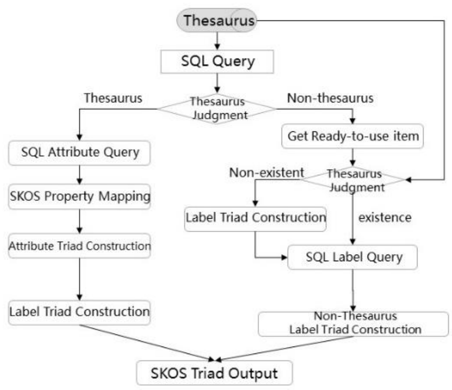

Figure 3. SKOS conversion technology flow chart.

In converting the Mapping Science Thesaurus and the Geoscience Science Thesaurus to SKOS, each thesaurus will be transformed into a concept of SKOS. As SKOS concepts, unique identifiers (URIs) are required elements to uniquely identify instances of concepts (skos: Concept). During the construction practice of Linked Data, the application of HTTPURLs is promoted to identify resources. Thesaurus of CAT has a stable and unique internal system number (term-code). Therefore, the term-code will be used as part of the HTTPURLs dynamically generated template "http://lod.aginfra.cn/cat/concept/{term-code}" when its thesaurus conversion is conceptualized to ensure that the unique and stable HTTPURLs to identify and parse the individual concepts in both tables.

SKOS provides label attributes such as skos:prefLabel and skos:altLabel to associate preferred and substituted natural language labels with specific concepts. In this paper, the Chinese strings and the corresponding English strings of the thesaurus in the two tables are mapped as skos:prefLabel with language markers, while the non- thesaurus of their "alternatives" is expressed as skos:altLabel with the same language markers. The semantic relationships in the two tables mainly include the types of "use, generation, genus, division, and reference". "In this paper, we use skos:broker, skos:narrower, and skos:related to translate the three relationships of "genus, subdivision, and reference" respectively. Next, we take the thesaurus "remote sensing image" from the Mapping Science Thesaurus as an example, and express the related thesaurus and inter-word relationships using the SKOS-based descriptive model. Figure 4 shows the information and inter-word relationships of "remote sensing images" extracted from the Mapping Science Thesaurus.

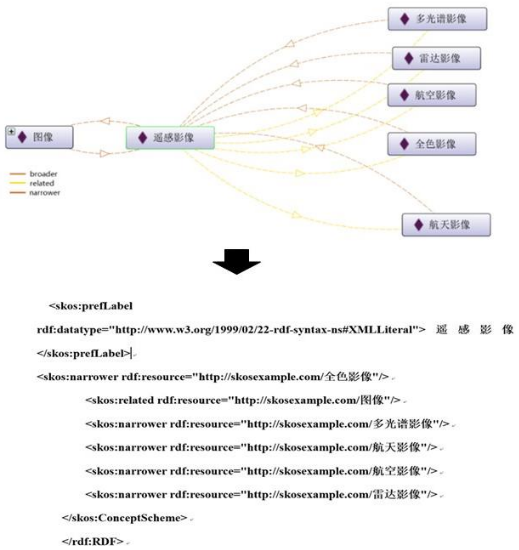

Figure 4. SKOS descriptive model of "remote sensing images".

3.2 Geographical entity extraction

Geographic named entity recognition, i.e., identifying geographical names in the geographic information-related professional literature and transforming them into structured GIS data, is the basis and key for mining geographic information from literature and extracting research areas in literature. By means of geographic parsing and geocoding based on natural language processing and with the help of the constructed ontology of location names, the location names of administrative divisions above the county level in China are identified from the literature and mapped to some space on the earth's surface that can be expressed using geometric types such as polygons, so as to assign geographical coordinates and geographical semantics to them. The location named entity identification (research area extraction) mainly includes the processes of natural language processing, location names ontology relationship database generation, geographic parsing, location names disambiguation, and geographic entity identification, and its architecture is shown in Figure 5 below.

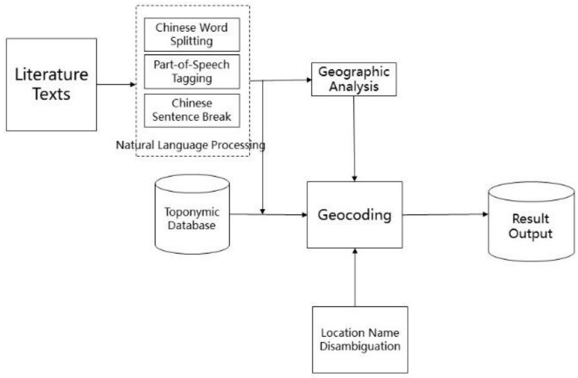

Figure 5. Geographical entity extraction technology roadmap.

Geographic entity recognition utilizes various resources and natural language processing methods provided by the HanLP natural language processing package, including corpus collection, Chinese word separation, and semantic annotation. The main task of natural language processing is to pre-process the abstract text in the input documents, including Chinese word separation, Chinese lexical annotation, and Chinese sentence breaking. The language used in Chinese text is Chinese, and the major difference between Chinese and English in form is that there is no clear separator between the words constituting a sentence, the sentences are separated by punctuation marks, and a sentence is a continuous string of Chinese characters. Therefore, word separation is the starting point and foundation of geographically named entity recognition. The system uses HanLP to complete the Chinese word separation, and at the same time to complete the annotation of lexical properties.

China's administrative divisions above the county level (including the county level) are divided into three major categories: provincial, prefectural, and county level. Among them, provincial-level administrative divisions are divided into provinces, municipalities directly under the central government, autonomous regions, and special administrative regions. The prefecture-level administrative divisions are divided into autonomous prefectures, regions, leagues, and prefecture-level cities. County-level administrative divisions are divided into county-level cities, counties, autonomous counties, flags, autonomous banners, special districts, forest districts, and municipal districts. Based on this, the ontology and relationship database of geographical names constituted the knowledge base of the administrative division of China in the field of geographical names, which becomes the knowledge source of Chinese geographical names identification and disambiguation. The natural language processing is to perform word separation and lexical annotation on the literature abstracts and to match the obtained location names with the relationship database of the location name ontology to obtain the administrative division attribution of each location name. There are two types of ambiguities that widely existed in Chinese location names: geo/non-geo ambiguity and geo/geo ambiguity. The geo/non-geo ambiguity arises when a location name has a non-geographic meaning. In layman's terms, geo/non-geo ambiguity is caused by the fact that a location name is the same as a common noun. geo/geo ambiguity is mainly generated by the use of the same location name in multiple geographical locations, i.e., the same name in different places. For the same location name may have multiple administrative affiliations, the administrative affiliation is determined according to the administrative level in order of background knowledge; for the case of multiple affiliations of the same location name, the disambiguation is based on the administrative affiliation of other high levels within the article.

3.3 Spatial data type extraction

Spatial data named entity recognition, that is, finding specific types of names in geographic information-related professional literature such as remote sensing images, topographic maps, vector maps, etc. This paper proposes a spatial data named entity extraction method based on CRF (Conditional Random Field) combined with rule method, CRF method has been applied to various fields of natural language processing, such as word separation, lexical annotation, named entity recognition, because natural language processing is not exactly a stochastic process, using statistical-based methods for spatial data named entity recognition alone will make the state search space very large and make it difficult to achieve the desired recognition accuracy and recall rate, which can then be corrected with the help of manually set rules for incorrect recognition results to improve the recall rate of spatial data named entity recognition. The technical flow chart is shown in Figure 6.

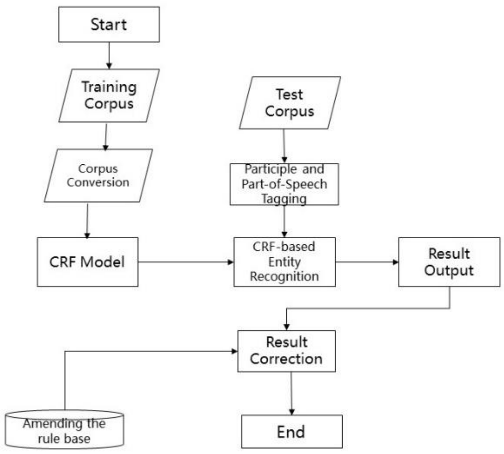

Figure 6. Flow chart of named entity recognition based on CRF and rule space data.

Step 1 Firstly, 50,000 training utterances and 1,000 non-overlapping test corpus are randomly selected from the literature abstracts of cartography and GIS disciplines from 1995 ~ 2017; the training corpus is annotated and transformed, and the transformed corpus is trained using the CRF model to generate the model parameters; Step 2 The open source HanLP word separation software is used to split the test corpus and lexical annotation, and the CRF model obtained in the previous step is used for the recognition of geographically named entities, and the word form and lexical annotation sequences are converted into the annotation set sequences defined in this paper.

The CRF algorithm alone cannot fully and accurately identify all named entities in spatial data, and the analysis of the reasons for non-identification is caused by sparse data or obscure features. For example, the "aerial data" that appears in the literature cannot be recognized, but it is actually an abbreviation for "aerial image map", but since this did not occur in the training, "aerial data" was not recognized as "aerial image" in the test. "was not recognized as "aerial image" in the test. For the above unrecognized cases, we design manual rules to correct the recognition results of spatial data naming entities. Firstly, we construct a dictionary of spatial data naming abbreviations, such as "ortho", "aerial", "elevation", etc., as well as backward and forward collocation words (ConjWord), such as "based on", "use", "of", etc. Then define the candidate string with recall containing spatial data named entity: WfSLOCWh, where SLOC = S1S2...Sn denotes the candidate location name, Si denotes the word in the candidate spatial data named entity, Wh denotes the post collocation word of the spatial data named entity, and Wh denotes the post collocation word of the location name, and first need to make rule correction from the sentence that needs to be Find such a token, and then judge SLOC using the rules defined below. The following definitions are designed, spatial data named entity abbreviation (SingleLoc), spatial data named entity suffix (LOC-E), ∈ for "belongs to", ‖ for "or", & & for "and ". The meaning of the rule is that when the proposition of the rule is true, the word satisfying the rule is judged to be geographically named. The design rule is as follows:

(Si ∈ SingleWord' & & Sn ∈ LOC - E) & & WfWh ∈ ConjWord) rule is mainly for the above example, where "based on" and "of" are the pre and post-collocations of aerial film and data respectively. The "aerial film" is the abbreviation of "aerial image" and "data" is the entity suffix, and then the boundaries of the named entities of spatial data are determined by the pre and post-collocations. Finally, identify "aerial data" as "aerial image" spatial data naming entity.

4. KNOWLEDGE SERVICES BASED ON GEOGRAPHICAL KNOWLEDGE GRAPH

In general, the current research on knowledge service applications based on geographic knowledge graphs is still in the initial stage, and with the rapid development of "Internet+", big data, and cloud computing, the research on knowledge services represented by knowledge graphs and knowledge center is flourishing, which provides a reference for knowledge service applications based on geographic knowledge graph (Chen et al.2019). The geographic knowledge graph studied in this paper has been integrated into the China Engineering Science and Technology Knowledge Center-Geographic Information Knowledge Service Platform, and representative knowledge service applications including knowledge graph visualization, data recommendation service, and the intelligent question and answer system have been developed and completed.

4.1 Knowledge graph visualization applications

Knowledge graph visualization is the most basic knowledge service application. Compared with the traditional relational database, the knowledge graph stored with graph data structure can view and analyze the entities and relationships of the knowledge graph more intuitively. The effect of the geographic knowledge graph constructed in this paper (partly) using the neo4j plug-in for visualization is shown in Figure 1. The scientific research literature entity nodes themselves exist independently, but these nodes can be associated with similar nodes of another scientific research literature entity node through their associated concepts, research areas, or spatial data possible so that the independent scientific research literature nodes are indirectly associated, and the whole knowledge graph constitutes a knowledge network. These relationships can be clearly observed and analyzed in Figure 7.

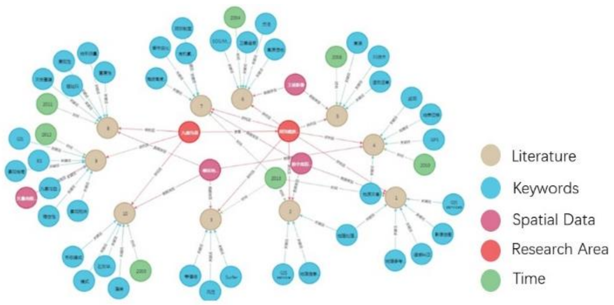

Figure 7. Knowledge graph visualization.

4.2 Data recommendation system applications

Since the disaster knowledge graph is constructed from disaster concept relationships, research areas are extracted from relevant literature, spatial data types, and the relationships between them. Therefore, the disaster knowledge graph can be applied to disaster research recommendation systems. The recommendation system can recommend the area where the disaster occurs more frequently as the user's research area, and provide the user with the common spatial data types of the research area based on the research data types selected by other scholars in the original scientific literature so that the user can recommend the spatial data he needs. In addition, since knowledge graphs can perform knowledge inference, new research data can be mined through relevant rules to achieve the update of recommended data. Figure 8 shows the main distribution provinces of "mudslide" in China recommended by the platform based on the knowledge graph when users search for "mudslide", and Figure 9 shows the "mudslide" recommended by the platform based on the knowledge graph, and then overlaying the magnitude of these data to the main distribution provinces.

Figure 8. Recommended applications for the research area.

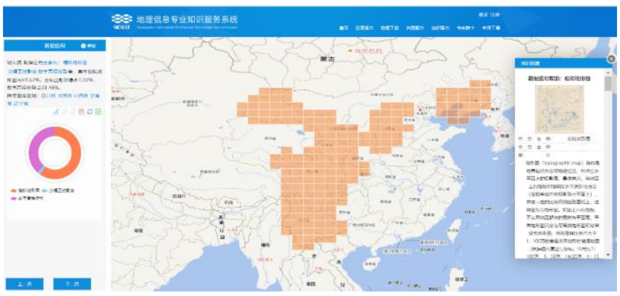

Figure 9. Recommended applications for spatial data types.

4.3 Intelligent question-answer system applications

The intelligent question-answer system based on the disaster knowledge graph is an important direction of this knowledge graph application. This question-answer system can accept natural language questions related to disasters from users, make statement queries in the established disaster knowledge graph, and finally return the corresponding answers to users. Natural language is closer to human communication habits in the form of interaction, and because the geographic knowledge graph has the characteristics of structured and correlated, geographic knowledge graph has the advantages of richer semantic expression, more accurate data content, and more efficient retrieval compared with pure textual information and structured databases, etc (Jiang et al.2018). Based on this, the question-answer system based on the disaster knowledge graph can better answer questions in related fields.

5. SUMMARY AND PROSPECT

In this paper, we analyze the concept of disaster knowledge graph and the current research status, and we propose a method and process for building the disaster knowledge graph with multi-source heterogeneous data, the core work of this knowledge graph is to construct the relationship between disaster concept- research area-spatial data. The disaster concept vocabulary and relationships are mainly transformed by the existing structured geographic thesaurus, and the rich semantic relationships (including the use of surrogate sub-references) provide rich semantic relationships of domain concepts for the entire knowledge graph, but the domain concept vocabulary is constantly growing and changing, so how to automatically expand the existing narrative concepts by the vocabulary embedded in the massive scientific research literature is also an important research direction at present. The research area and spatial data type are mainly obtained through knowledge extraction methods, where the research area is not only the spatialization of the literature, but also provides the spatial dimensional association relationship for the whole knowledge graph, and can be a crucial foundation for the spatial visualization application of the knowledge graph. The spatial data extraction is mainly used to establish the association relationship between the disaster field and the mapping and geographic information field, which can provide data recommendation service to the disaster research workers based on this relationship, and to the mapping and geographic information field, which can clarify the main applications of the data produced by itself, and can actively carry out spatial data recommendation for disaster research. The disaster knowledge graph has a wide range of applications. In addition to the visualization and data recommendation applications introduced in this paper, it can also be used for research on relationship prediction and intelligent decision analysis and has a wide range of scientific and commercial application scenarios.

ACKNOWLEDGEMENTS

This research was supported by Scientific Research Projects of Beijing Municipal Education Commission—General Projects of Science and Technology Program (Surface Projects) (KM202110016003); Beijing Key Laboratory of Urban Spatial Information Engineering (No. 2020202); The National Natural Science Foundation (NSFC) of China (Key Project #41930650); and the Fundamental Research Funds for Beijing University of Civil Engineering and Architecture (X20046).

REFERENCES

Buscaldi, D., Rosso, P., 2008. A conceptual density-based approach for the disambiguation of toponyms. *International Journal of Geographical Information Science.*, 22(3), 301-313.

Chen, J., Liu, W.Z., Wu, H., Li, Z.L., Zhao, Y., Zhang, L., 2019.Basic Issues and Research Agenda of Geospatial Knowledge Service (in Chinese). *Geomatics and Information Science of Wuhan University (01)*.,38-47. DOI:10.13203/j. whugis20180441.

Davies, C., 2013. September. Reading geography between the lines: extracting local place knowledge from text. *In International Conference on Spatial Information Theory* (pp. 320-337). Springer, Cham.

Dojchinovski, M., Kliegr, T., 2013. Entityclassifier. eu: real-time classification of entities in text with Wikipedia. *In Joint European Conference on Machine Learning and Knowledge Discovery in Databases* (pp. 654-658). Springer, Berlin, Heidelberg.

Guo, X.Y., He, T.T., 2015. Survey about Research on Information Extraction (in Chinese). *Computer Science (02)*,14-17+38.

Hua, B.L., Zhang, X.M., 2010. Describing Character and Trend of Knowledge Extraction Based on Discussion on Concepts Relative with Knowledge Extraction (in Chinese). *Information Science (02)*., 311-315.

Jiang, B.C., Wan, G., Xu, J., Li, F., Wen, H.Q., 2018.GeoGgrphic Knowledge Graph Buliding Extracted from Mulit-sourced Heterogeneous Data (in Chinese). *Acta Geodaetica et Cartographica Sinica (08)*.,1051-1061.

Lieberman, M. D., Samet, H., 2011. Multifaceted toponym recognition for streaming news. *In Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval* (pp. 843-852).

Li, G.F., Zhang, P., 2018. A Web knowledge extraction model based on agricultural ontologies (in Chinese). *Jiangsu Agricultural Sciences (04)*,201-205. DOI: 10.15889/j.issn.1002-1302.2018.04.051.

Li, H.D., 2018. Research on Medical Domain Knowledge Extraction Methods (in Chinese). Harbin Institute of Technology.

Lu, F., Yu, L., Qiu, P.Y., 2017. On Geographic Knowledge Graph (in Chinese). *Journal of Geo-information Science (06)*,723-734.

Marrero, M., Urbano, J., Sánchez-Cuadrado, S., Morato, J., Gómez-Berbís, J. M., 2013. Named entity recognition: fallacies, challenges and opportunities. *Computer Standards & Interfaces.*, 35(5), 482-489.

Ni, Y., Zhang, L., Qiu, Z., Wang, C. 2010., Enhancing the open-domain classification of named entity using linked open data. *In International Semantic Web Conference* (pp. 566-581). Springer, Berlin, Heidelberg.

Qiu, Q.Y., Xue, C., Feng, P.E., Deng, K., 2012. Novelty Evaluation Method of Patent Design Knowledge and Its Application in Creative Design (in Chinese). *Journal of Mechanical Engineering (11)*,39-45.

Suchanek, F. M., Kasneci, G., Weikum, G., 2008. Yago: A large ontology from wikipedia and wordnet. *Journal of Web Semantics*., 6(3), 203-217.

Twaroch, F. A., Smart, P. D., Jones, C. B., 2008. Mining the web to detect place names. *In Proceedings of the 5th Workshop on Geographic Information Retrieval* (pp. 43-44).

Wen, W., Wu, S.J., Cai, R.C., Hao, Z.F., 2018. Type Extraction and Labelling of Knowledge Entities in the Field of Professional Literature (in Chinese). *Journal of Chinese Information Processing (01)*,102-115.

Xu, H.J., Zeng, W., Zhang, Y.L., 2018. Paper-Patent Topic Linkage Evolution Analysis Method Based on Word2vec (in Chinese). Journal of Intelligence (12),36-42.

Yang, Y.J., Xu, B., Hu, J.W., Tong, M.H., Zhang, P., Zheng, L., 2018. Accurate and Efficient Method for Constructing Domain Knowledge Graph (in Chinese). *Journal of Software (10)*,2931-2947. DOI: 10.13328/j.cnki.jos.005552.

Yu, S., He, T., Glass, J.R., 2020a. Constructing a Knowledge Graph from Unstructured Documents without External Alignment.

Zhai, D.W., 2017. Research and Application of Patent Information Extraction and Topic Mining (in Chinese). Beijing University of Technology.