Intelligent Human Systems Integration (IHSI 2023), Vol. 69, 2023, 302–311
https://doi.org/10.54941/ahfe1002849

An Intelligent Retrieval Method of Building Fire Safety Knowledge Based on Knowledge Graph

Fengyang Sun and Beibei Sun

Southeast University, Nanjing, Jiangsu Province, China

ABSTRACT

Aiming at the difficulty of searching standards and specifications due to the huge and fragmented data in the field of fire safety in China, an intelligent retrieval method of building fire safety knowledge based on knowledge graph is proposed. First, ontology is used to construct conceptual schemas from top down, and knowledge is extracted using rule templates and stored in the Neo4j graph database to complete the construction of knowledge graph. Then, on the basis of the knowledge graph, BERT-BiLSTM-CRF model and BERT classifier are used to process complex questions with multiple constraints, so as to extract key entities in the question and identify query intention. Finally, according to the key entities and query intention, an algorithm is used to generate a Cypher query statement, which is used to obtain the answer in Neo4j. The intelligent retrieval method based on knowledge graph standardizes the building fire safety knowledge, solves the problem of scattered distribution and greatly improve the efficiency of knowledge retrieval.

Keywords: Intelligent retrieval, Knowledge graph, Building fire safety, Deep learning

INTRODUCTION

Fire is one of the most frequent and common disasters that threaten public safety and social development. Faced with the challenges and threats brought by fire, comprehensive prevention of fire risks has become an important safety requirement of the society, and it is also a manifestation of the civilization level of modern society and the ability of social public management. However, there are many irregularities in the national standards on which fire risk assessment relies, mainly manifested as scattered documents and irregular names, which are difficult to meet the requirements of modern fire safety. Based on knowledge graph and deep learning, this paper studies the intelligent retrieval of knowledge in the field of building fire safety, which can provide theoretical support for fire risk assessment and promote the construction of intelligent fire protection.

As an emerging technology, knowledge graph has been widely researched and developed since it was formally proposed by Google in 2012. Knowledge graph based on the thought of semantic network, essentially is a semantic knowledge base with a directed graph structure, which is used to describe concepts and their relationships in the physical world in a symbolic form (Liu et al. 2016). It expresses the information of the Internet into a form more consistent with human cognition, and provides a better ability to organize, manage and understand the massive information on the Internet. Knowledge graph plays an important role in information retrieval, recommendation system, decision analysis and other aspects, and is widely used in medical treatment (Li et al. 2020), finance (Cheng et al. 2020) and other fields. In the field of fire safety, Mu et al. (2019) designed and developed an intelligent drawing review system based on BIM, the system introduced the knowledge graph to automatically review building models for fire protection. To a certain extent, the research provides data and framework basis for the construction of knowledge graphs in the field of building fire protection, but the research in this field is still in its infancy, there is a lack of mature results in both data extraction and knowledge ontology construction. Moreover, the data used for knowledge graph construction is mainly structured data such as BIM, ignoring the widely used unstructured data such as national and industry standards.

As the core application direction of knowledge graph, intelligent retrieval uses natural language processing (NLP) technology to analyse the semantic meaning of questions, identify the user's intention, and feedback the answer after reasoning calculation. Wang et al. (2021) proposed a BIM data-oriented multi-scale building fire information retrieval method, which combined BIM and knowledge graph to realize multi-scale information retrieval in the field of building fire protection. Zhang et al. (2021) realized intelligent retrieval of agricultural knowledge by constructing a knowledge graph in the field of agriculture. At present, most of the information retrievals in the vertical domain are based on rule templates, which are difficult to effectively use the information in the knowledge graph and cannot deal with the complex problems with multiple constraints.

In this paper, the national standards in the field of building fire safety are used as the data source to construct the knowledge graph. Through the knowledge representation of the graph, the huge standards and norms are effectively organized and managed to solve the problems of scattered distribution and non-standard representation. With the support of knowledge graph, deep learning methods are used to extract key entities and identify intentions for complex multi-constraint questions. Finally, answers are obtained by using dynamically generated query statements.

CONSTRUCTION OF KNOWLEDGE GRAPH

The construction of knowledge graph is the premise of intelligent retrieval and other upper-level applications. It fundamentally determines the quality of upper-level applications and is the most basic and important part of the whole system construction. Knowledge graph can be constructed in two ways: top-down and bottom-up (Xu et al. 2016). Since the knowledge structure in the field of building fire safety is hierarchical and mostly comes from structured data, this paper adopts the top-down approach to complete the construction of knowledge graph, that is, the pattern layer is constructed first, and then the knowledge is extracted. The construction process mainly includes three steps:

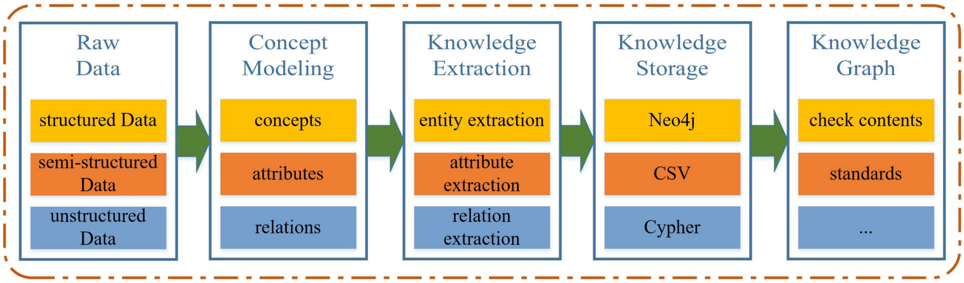

Figure 1: Construction process of knowledge graph.

concept modeling, knowledge extraction and knowledge storage, which is shown in Figure 1.

In the first step, ontology is used to construct the schema layer of knowledge graph, and conceptualized information is extracted. In the second step, according to the conceptualized information extracted, NLP methods are used to extract knowledge, including the extraction of entities, attributes and relations. In the third step, the knowledge is stored in a graph database and the construction is finished.

This paper takes Fire Safety Assessment (2021) and Code for Fire Protection (2018) as data sources to construct knowledge graphs. Firstly, the conceptual information of check items, check contents, check objects and corresponding standards in Fire Safety Assessment is summarized, and the schema layer is established by using ontology, as shown in Figure 2. Secondly, because the contents of Fire Safety Assessment and Code for Fire Protection are mainly organized in the form of numbered paragraphs and tables, the DOCX package and RE package in Python are imported to process the documents according to the specified rules to extract knowledge. At the same time, the extracted knowledge is saved into a CSV file by using the PANDAS package. Finally, the query language Cypher is used to import the knowledge from CSV files into a Neo4j database. Figure 3 shows a part of the constructed knowledge graph, which contains relationships between check contents and standards, and constraint relationships in Code for Fire Protection.

REALIZATION OF INTELLIGENT RETRIEVAL

For complex questions in which there are multiple conditional entities to constrain the target entity, a classification model based on Bert (Bidirectional Encoder Representation from Transformers) is used to classify questions to identify query intentions and a deep learning model Bert-BilSTM-CRF (Bert - Bidirectional Long Short-Term Memory - Conditional Random Field) is used to recognize the named entity in the question to extract the constraints. Finally, the Cypher query statement is constructed dynamically by combining the query intention and all constraints, so as to query the answer in the graph database.

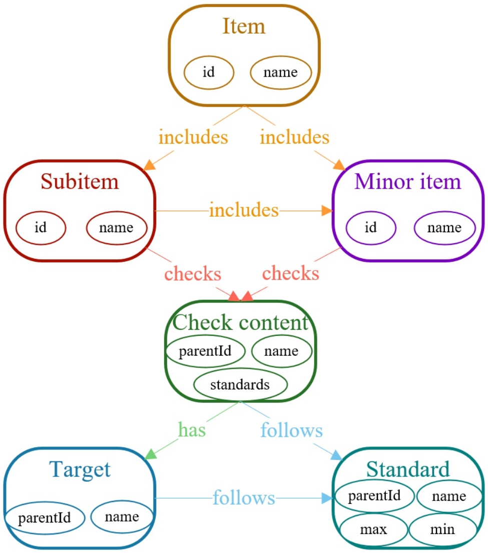

Figure 2: A part of knowledge graph schema layer.

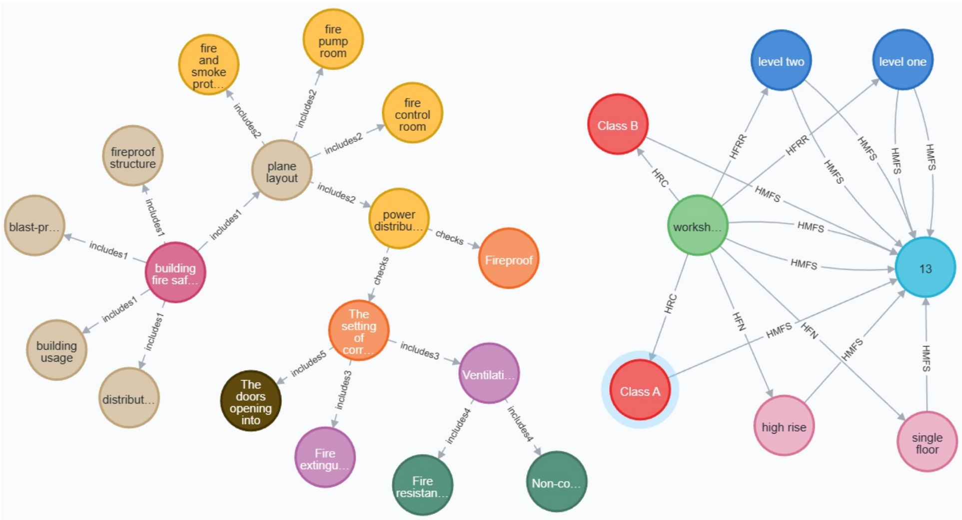

Figure 3: A part of the knowledge graph.

Question Classification Based on Bert

BERT is a language model pretrained with large-scale corpus (Devlin et al. 2019), which is based on two-way Transformer encoder (Vaswani et al. 2017). It can greatly reduce the training time, and has a strong semantic representation ability and a feature extraction ability, which can better identify the intention of a complex question. Therefore, this paper selects Bert-Base-Chinese model, which is pretrained with Chinese corpus, combined with a full connection layer to complete the classification task of multi-constraint complex questions and its training parameters are shown in Table 1. In this paper,

Table 1. Training parameters of classification model based on Bert.

Batch size	Padding size	Hidden size	Output size	Learning rate
16	50	768	12	1.5e-5

Table 2. Labels for named entity recognition.

Label	Meaning
O	Other
B-BU	The beginning of a building/location entity
I-BU	The inner of a building/location entity
E-BU	The end of a building/location entity
S-BU	A single building/location entity
…	…

the multi-constraint questions are divided into 12 classes, corresponding to fire separation, combustibility, fire resistance and so on.

Named Entity Recognition Based on Bert-Bilstm-CRF

Named entity recognition is the basis of many natural language processing applications, such as entity extraction, intelligent question answering, etc. It aims to automatically identify the required entities from unstructured textual data and label them, such as people's names and place names (Deng et al. 2021). Main NER methods can be divided into three categories: rule and dictionary based methods, statistical machine learning based methods and deep learning based methods (Li et al. 2022). Rule and dictionary based methods require manual construction of rule templates and recognize entities through matching (Shaalan & Raza, 2009). Methods based on traditional machine learning use feature templates and machine learning models for entity recognition. The commonly used models include Hidden Markov Model (HMM) (Zhang et al. 2004), Maximum Entropy Model (MEM) (Zhang et al. 2008) and Conditional Random Field (CRF) (Yan et al. 2014). Deep learning based methods use deep neural networks to automatically extract features and complete entity recognition by sequence labeling. Commonly used neural networks include ID-CNNS (Strubell et al. 2017), BilSTM-CRF (Huang et al. 2015) and Bert-BilSTM-CRF (Yin et al. 2020), etc. Deep learning based methods can achieve high accuracy without complex feature engineering, so the Bert-BilSTM-CRF model, which is widely used, has high accuracy and short training time, is selected in this paper to complete the named entity recognition task.

BIOES annotation system is used to label the questions. B denotes the beginning of an entity, I denotes the inner of an entity, E denotes the end of an entity, O denotes "other", and S denotes a single entity. In this paper, there are 69 kinds of labels in the named entity recognition task. Table 2 shows some labels and their meanings, and Figure 4 shows some examples of sequence labeling.

sentence	What	is	the	combustion	performance	of	the	load	bearing	wall
labels	O	O	O	O	O	O	O	B-BC	I-BC	E-BC
sentence	in	a	level	two	warehouse					
labels	O	O	B-FRR	E-FRR	S-BU					
sentence	What	is	the	maximum	number	of	floors	of	a	Class
labels	O	O	O	O	O	O	O	O	O	B-RC
sentence	C	and	level	one	workshop					
labels	E-RC	O	B-FRR	E-FRR	S-BU					
sentence	What	is	the	fire	separation	between	a	Class	A	,
labels	O	O	O	O	O	O	O	B-RC	E-RC	O
sentence	single	floor	,	level	one	warehouse	and	a	Class	B
labels	B-FN	E-FN	O	B-FRR	E-FRR	S-BU	O	O	B-RC	E-RC
sentence	,	high	rise	,	level	two	workshop			
labels	O	B-FN	E-FN	O	B-FRR	E-FRR	S-BU			
sentence	What	is	the	minimum	net	width	of	the	evacuation	door
labels	O	O	O	O	O	O	O	O	B-EZ	E-EZ
sentence	of	a	high	rise	medical	building				
labels	O	O	B-FN	E-FN	B-BU	E-BU				

Figure 4: Examples of sequence labeling.

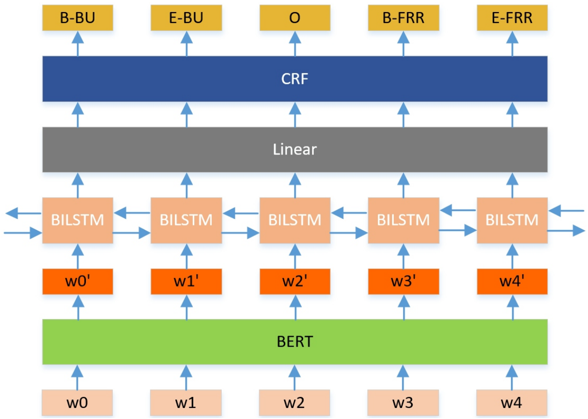

Figure 5: Structure of Bert-BiLSTM-CRF model.

The model structure of Bert-BILSTM-CRF is shown in Figure 5, which is mainly composed of four modules. The character sequence obtained from the question will be input into BERT to get the feature vector. BiLSTM will further extract the feature, and the linear layer will generate a preliminary label sequence. Finally, the CRF layer will constrain the dependencies between labels to get an optimal solution of the label sequence. Training parameters of the model are shown in Table 3.

The Bert-BilSTM-CRF model generates a label sequence after labelling the question, according to which the named entity in the question can be extracted.

Dynamic Generation of Cypher Query Statements

Firstly, the BERT classifier is used to classify the user's questions. According to the classification results, the target entity type of the question is determined. Then, the Bert-BilSTM-CRF model is invoked to recognize the entities, and the constraints in the question are extracted. After the target

Table 3. Training parameters of Bert-BiLSTM-CRF model.

Batch size	6
Padding size	50
Hidden size of Bert	768
Hidden size of LSTM	128
Output size	69
Learning rate of Bert	3e-5
Learning rate of BiLSTM and Linear	3e-4
Learning rate of CRF	3e-2
Dropout	0.3

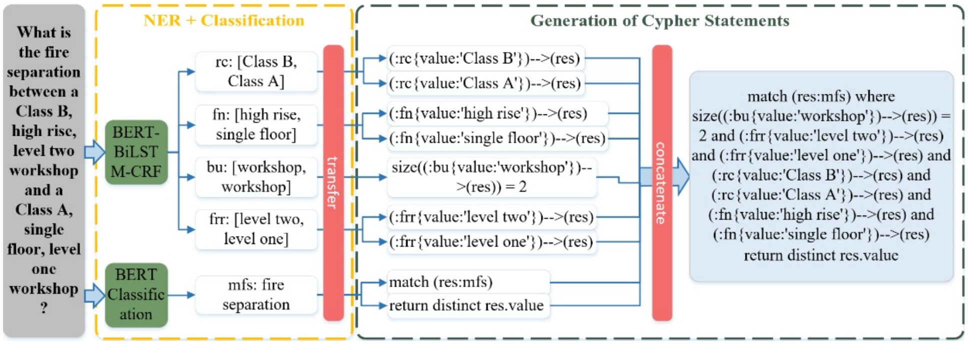

Figure 6: Process of Cypher statement generation.

entity type and constraints are obtained, an algorithm is invoked to generate a Cypher query statement. The algorithm specifies the entity type returned in the Cypher query statement according to the target entity type, and generates the Cypher statement fragments for each constraint. Finally, all Cypher fragments are concatenated into a complete query statement to retrieve the answer in Neo4j. The flow chart of dynamically generating query statements is shown in Figure 6.

INTELLIGENT RETRIEVAL SYSTEM

An intelligent retrieval system is built based on the proposed retrieval method. This system uses B/S architecture, Java language, Neo4j database to store the knowledge of building fire safety. SpringBoot is used in the back end, Vue and third-party libraries such as D3 and ElementUI are used in the front end. The BERT classification model and NER model run on the Python server, and the Java back-end communicates with Python services by establishing socket connections. The software architecture of the system is shown in Figure 7.

For the convenience of users, a question and answer window as shown in Figure 8 is designed, which supports manual input and voice input. The system provides a visual display of the knowledge graph, as shown in Figure 9. The left part visually displays all knowledge queried, and the right part displays information such as the number and name of the selected node. It also provides the functions of querying specified nodes and adjacent nodes,

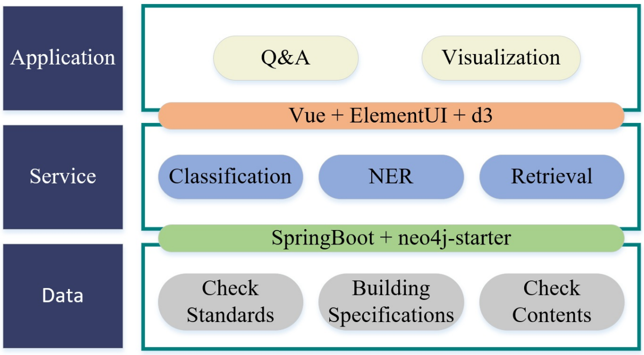

Figure 7: Software architecture.

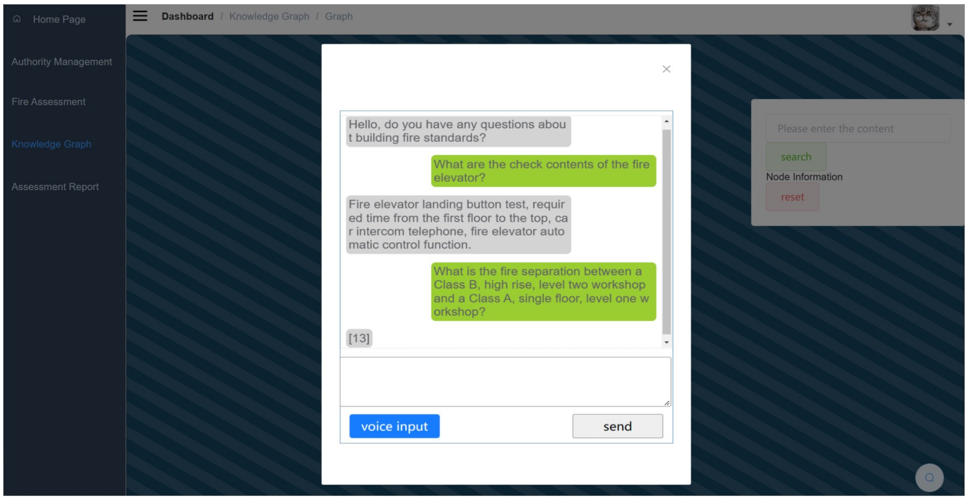

Figure 8: Question and answer window.

and resetting the graph. Through the system, users can find the knowledge in the field of building fire safety quickly and accurately.

CONCLUSION

In this paper, a knowledge graph in the field of building fire safety is constructed, an intelligent retrieval method for multi-constraint questions is proposed, and an intelligent retrieval system based on knowledge graph is established. Through the introduction of knowledge graph, norms in the field of building fire protection are organized, and problems of scattered and non-standard are solved. The intelligent retrieval system enables assessors to quickly and accurately find building standards and codes, which greatly improves the efficiency of fire risk assessment.

The knowledge graph and intelligent retrieval system still have some shortcomings and room for improvement, mainly including: (1) The data in the knowledge graph is not rich enough and should be supplemented by other data sources; (2) Since the amount of data used for training NER model in

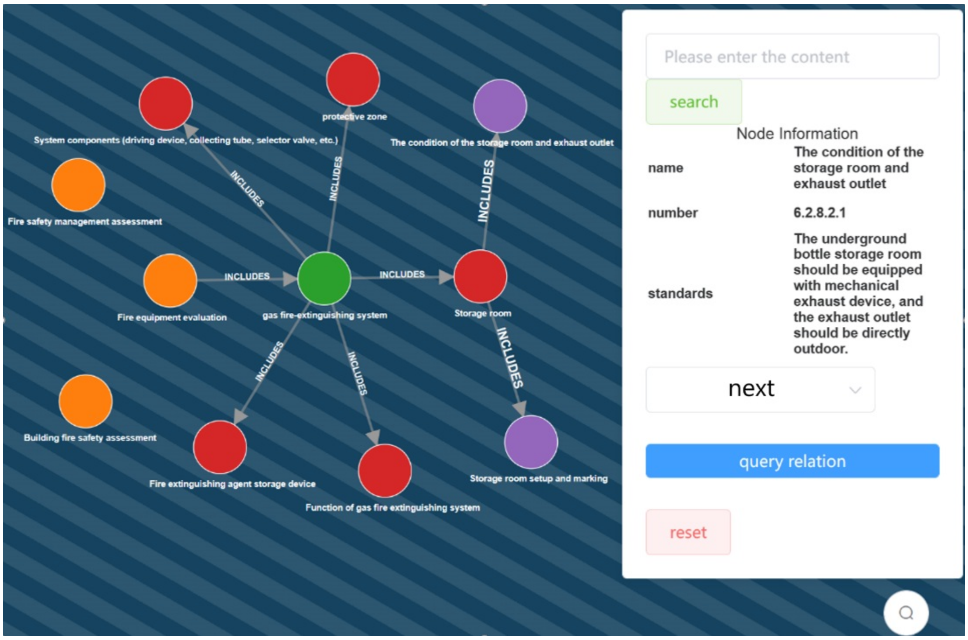

Figure 9: Visualization of knowledge graph.

this paper is not large enough, the recognition rate of those rarely mentioned entities is relatively low. Therefore, users' questions can be saved during the use of the retrieval system and added to the training data to improve the recognition rate.

ACKNOWLEDGMENT

I would like to acknowledge my tutor, who spent a lot of time reading my paper and helping me revise it carefully. At the same time, I am very grateful to my classmates, they give me a lot of suggestions in life and help me overcome many difficulties.

REFERENCES

Cheng, D. Yang, F. Wang, X. Zhang, Y and Zhang, L. (2020). "Knowledge graph-based event embedding framework for financial quantitative investments", Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, New York, NY.

Deng, Y. Wu, C. Wei, Y. Wan, Z and Huang, Z. (2021). A Survey on Named Entity Recognition Based on Deep Learning, Journal of Chinese Information Processing Volume 35 No.9.

Devlin, J. Chang, M. Lee, K and Toutanova, K. (2019). "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding", Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Stroudsburg, PA.

Huang, Z. Xu, W and Yu, K. (2015). Bidirectional LSTM-CRF models for sequence tagging, arXiv: 1508.01991.

Li, D. Luo, S. Zhang, X and Xu, F. (2022). Review on Named Entity Recognition, Journal of Frontiers of Computer Science and Technology Volume 9.

Li, L. Wang, P. Yan, J. et al. (2020). Real-world data medical knowledge graph: construction and applications, Artificial intelligence in medicine Volume 103.

Liu, Q. Li, Y. Duan, H. Liu, Y and Qin, Z. (2016). Knowledge Graph Construction Techniques, Journal of Computer Research and Development Volume 53 No.3.

Ministry of Emergency Management of the People's Republic of China. (2021). Fire Safety Assessment (XF/T 3005-2020). Beijing: Emergency Management Press.

Mu, L. Wang, J. Li, J and Zhou X. (2019). An intelligent review of fire drawings based on building information modelling and knowledge graph, Fire Science and Technology Volume 38 No.12.

Shaalan, K and Raza, H. (2009). NERA: Named entity recognition for Arabic, Journal of the American Society for Information Science and Technology Volume 60 No.8.

Strubell, E. Verga, P. Belanger, D and McCallum, A. (2017). "Fast and accurate entity recognition with iterated dilated convolutions", Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Stroudsburg, PA.

The Ministry of Public Security of the People's Republic of China. (2018). Code for Fire Protection (GB 50016-2014). Beijing: China Planning Press.

Vaswani, A. Shazeer, N. Parmar, N. et at. (2017). "Attention Is All You Need", Proceedings of the 2017 conference on neural information processing systems, Stroudsburg, PA.

Wang, J. Gao, X. Xiao, L. Li, J and Zhou, X. (2021). Research on multi scale building fire information retrieval method based on BIM, Fire Science and Technology Volume 40 No.4.

Xu, Z. Sheng, Y. He, L and Wang, Y. (2016). Review on knowledge graph techniques, Journal of University of Electronic Science and Technology of China Volume 45 No.4.

Yan, Y. Wen, D. Wang, Y and Wang, K. (2014). Named entity recognition in Chinese medical records based on cascaded conditional random field, Journal of Jilin University (Engineering and Technology Edition) Volume 44 No.6.

Yin, X. Zhao, H. Zhao, J. Yao, W and Huang, Z. (2020). Multi-neural network collaboration for Chinese military named entity recognition, Journal of Tsinghua University (Science and Technology) Volume 60 No.8.

Zhang, H. Chen, Q. Zhang, S. Zhang, Z. Yang, F and Li, X. (2021). Intelligent Retrieval Method of Agricultural Knowledge Based on Semantic Knowledge Graph, Transactions of the Chinese Society for Agricultural Machinery Volume 52 No.S0.

Zhang, J. Shen, D. Zhou, G. Su, J and Tan, C. (2004). Enhancing HMM-based biomedical named entity recognition by studying special phenomena, Journal of Biomedical Informatics Volume 37 No.6.

Zhang, Y. Xu, Z and Xue X. (2008). Fusion of multiple features for Chinese named entity recognition based on maximum entropy model, Journal of Computer Research and Development No.6.