International Journal on Semantic Web and Information Systems
Volume 19 • Issue 1

Research on the Construction and Application of Knowledge Graph in the Ceramic Field Based on Natural Language Processing

Yu Nie, Jingdezhen Ceramic University, China*
https://orcid.org/0000-0001-6549-4660

Na Huang, Jingdezhen Vocational University of Art, China

Junjie Peng, Jingdezhen Ceramic University, China

Guanghua Song, Jingdezhen Ceramic University, China

Yilai Zhang, Jingdezhen Ceramic University, China

Yongkang Peng, Jingdezhen Ceramic University, China

Chenglin Ni, Guangdong Songfa Ceramics Co., Ltd., China

ABSTRACT

There are problems of knowledge deficiency and effective unified expression of knowledge in the process of relevant knowledge data acquired by workers in the ceramic domain. In this study, the authors designed relevant experiments to construct ceramic field knowledge graphs to solve these problems. In the experiments of named entity recognition and relationship recognition, the authors compared the performance of several models in OwnThink and ceramics field datasets. The experimental results showed that the BiLSTM-CRF model is the best for named entity recognition and the TextCNN model is the best for relationship recognition in ceramics field datasets. Therefore, the first used the BiLSTM-CRF model to complete the naming entity recognition and then combined with the TextCNN model to complete the relationship recognition to construct the ceramic field knowledge graph. Then, they applied the constructed graph to the ceramic knowledge Q&A service to provide accurate data retrieval service for ceramic domain workers.

KEYWORDS

Ceramics Field, Entity Recognition, Knowledge Graph, Natural Language Processing, Relationship Recognition

INTRODUCTION

At present, the knowledge data obtained by most workers in the ceramic domain through the above methods are not only time-consuming and labor-intensive, but also may have the problem of knowledge loss. In addition, the acquired knowledge also lack of unified expression and effective organization,

DOI: 10.4018/IJSWIS.327352

*Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

which is difficult to be directly used in machine learning of data in the ceramic field. On the other hand, in the field of ceramic data, because the data are too scattered, there is a small sample problem, which also brings challenges to machine learning.

How to collect, organize, and manage this scattered knowledge in the field of ceramics and how to support the scientific research of ceramics are urgent problems to be solved.

It is well-known that, in 2012, Google proposed a new search function—knowledge graph-based search—aiming to improve search performance; the concept of knowledge graph has gradually become known to everyone. The knowledge graph explicitly precipitates and correlates domain knowledge, which can well solve the characteristics of scattered, complex, and isolated data in the domain (Xu et al., 2016). The question-answering system deals with natural language questions. The first thing to do is to identify the named entity of the question, and then retrieve the knowledge from the knowledge base and extract the corresponding answer. Knowledge graphs are used for knowledge retrieval and answer extraction due to their knowledge reasoning capabilities, which provide Q&A systems with a knowledge base that has a recommendation mechanism and higher accuracy. The two major links provide a knowledge base with a recommendation mechanism and higher precision for the question-answering system (Xia, 2016).

In recent years, knowledge graph technology has been applied in many fields and achieved remarkable results (Wu et al., 2021), such as smart finance, smart medical care, smart education, and smart e-commerce (Fu et al., 2021; Hang et al., 2019). However, in the ceramic field, the related research of knowledge atlas is still in the initial stage (Cao, 2021; Katiyar & Cardie, 2017). Besides, only a few researchers have carried out exploration on ceramic knowledge extraction (Li, 2020; Zheng, Hao et al., 2017; Zheng, Wang et al.,2017), carried out the extraction research of general type knowledge such as ceramic entity and relationship, rarely involved in the acquisition method of knowledge with ceramic field characteristics such as process flow, and almost no research on establishing ceramic knowledge atlas and applying it to ceramic machine learning.

The application of knowledge graph technology in the ceramic domain can effectively utilize the advantages of knowledge graphs in domain knowledge learning, organization, and reasoning, which is helpful to solve the challenging problems of machine learning in the ceramic domain.

In the process of knowledge graph construction, entity and relationship extraction is the foundation, entity alignment is the guarantee of knowledge fusion, and knowledge graph completion is the core to improve the quality of knowledge graph. In the process of entity and relationship extraction, semantic analysis and understanding of content information cannot be separated from the support of a semantic knowledge base (Li & Ji, 2014; Wang et al., 2016), and the machine can understand human language through a semantic knowledge base, thus becoming more intelligent. However, most of the traditional semantic knowledge bases are oriented to general fields, which cannot meet the knowledge requirements of natural language processing systems in specific fields (Guo et al., 2018; Zeng et al., 2018). Therefore, the authors will solve the related problems of ceramics by constructing knowledge graphs of the ceramic domain.

Thus, in this study, the authors applied knowledge graph technology to ceramics by collecting ceramics-related data on the Internet, effectively using the advantages of knowledge graph in domain knowledge learning, organization, and inference, which can help solve the challenging problem of machine learning in ceramics, and at the same time provide more accurate data retrieval services for ceramics field workers.

RELATED WORK

The vigorous development of knowledge graphs has greatly promoted the practical application of knowledge tests, and breakthrough progress has been made in this field (Liu et al. 2016; Xu et al., 2016). From the perspective of analysis methods, the research methods based on knowledge graphs can be divided into two types: The semantic-based method and the retrieval-based method. The semantic-based method mainly transforms the question into a more logical form and then uses structured query statements to obtain appropriate answers in the knowledge graph. Zettlemoyer and Collins (2012) and Kwiatkowski (2010) adopted the corpus with logical formal annotations and supervised training. Liang et al. (2011) proposed to replace regular expressions with structured queries based on dependency composition semantics. The latter is usually more focused on efficiently extracting the features of questions or answers and the order of correct answers. In response to this problem, Zhang et al. (2015) designed and implemented an aviation-oriented question answering system centered on the ontology triplet. the expected effect. Yao and Durme (2014) used dependency analysis to characterize the question sentence of the question to obtain a relationship graph and locate the relationship and entity within a certain range. By comparing the resulting subject graph and the graph generated by the question, all nodes are sorted and manipulated along the graph to obtain the correct answer.

Currently, two main ways are available to construct knowledge graphs. One is top-down, which directly extracts ontology and schema information from external third-party databases to construct a knowledge base, which is a simple and effective method to try to construct a knowledge graphs rudiment, but the accuracy is insufficient. The second is a bottom-up method, which uses artificial intelligence technology to realize the automatic acquisition and processing of knowledge to construct a knowledge base according to mass data on the Internet. Because of the rapid development of artificial intelligence technology in recent years, this kind of method has become the first choice for constructing knowledge graphs.

For the bottom-up construction method, after data acquisition and preprocessing, according to the acquisition logic, a series of construction techniques are used to realize the tasks of knowledge extraction, knowledge fusion, and knowledge processing reasoning. The process is a continuous iterative update process. Firstly, entities, attributes, and relationships among entities are found in the given multisource data to form knowledge representation. Subsequently, because of the multiheterogeneous sources of knowledge, it is necessary to use knowledge fusion technology to eliminate the contradiction and redundancy of knowledge. Thirdly, reasoning the fused knowledge, predicting the new knowledge, and expanding the existing knowledge to ensure the integrity of the knowledge graph and expand its scale.

The entity extraction and the relationship extraction are used to obtain factual knowledge such as entities and relationships thereof from multisource data, to form knowledge triples $G \in (E, R, S)$ required by the knowledge graph, where E represents a set of entities in the knowledge base, R represents a set of relationships in the knowledge base, and S $E \times R \times E$ represents a set of triples in the knowledge base, which specifies the number of entities, the number of types of relationships, and the number of triples formed. Entity alignment is used to eliminate synonymous heterogeneous entities in the multisource heterogeneous knowledge base and remove data redundancy to complete knowledge fusion. Knowledge graph completion is to improve the quality of knowledge graph and expand the scale of knowledge graph by using existing knowledge reasoning or data mining implicit fact knowledge.

As a new knowledge representation method, a knowledge graph belongs to the category of the Semantic Web, which is used to describe various entities existing in the real world and the relationship between entities with abstract concepts (Wang et al., 2018; Zhu et al., 2019,). It can be divided into two categories, according to the coverage. One is universal knowledge graph, and the typical universal knowledge graphs include Open Cyc, Word Net, and YAGO. The other is the domain knowledge graph. Typical domain knowledge graphs include Palantir, Plant Data, and Ace KG. The general knowledge graph emphasizes entity quantity and domain coverage and does not require very high accuracy. It is mainly used in search engines to recommend more retrieval results for users in the form of visual graph association according to user input. Domain knowledge graphs focus on the relationship between entities in specific fields or industries and require high accuracy. The knowledge system constructed by the domain knowledge graphs is characterized by strong domain pertinence and professionalism (Ruan et al., 2016); this knowledge system is mainly used to assist complex analysis or decision support in industries. At present, it has been applied in patent information, venture capital enterprises, and other fields. In contrast, a domain knowledge graph has more important significance for practical production and life, so the main research object of this topic is the construction of a ceramic domain knowledge graphs.

DATA SOURCES

In this research, the authors used a variety of search engines (including Baidu, Bing, and Sogou), and, after retrieving various search results, finally selected the data of the Liao Liao Ting Web site as the ceramic field dataset. Liao Liao Ting, as the Jingdezhen ceramic art Web site, enjoys a high reputation in the contemporary ceramic collection circle. The ceramic dataset the authors constructed in this study includes three aspects: Ceramic technology, ceramic art, and ceramic history.

CONSTRUCTION OF KNOWLEDGE GRAPH IN THE FIELD OF CERAMICS

Figure 1 shows the construction process of knowledge graph and corpus.

Data Acquisition in the Field of Ceramics

In this project, the authors used a variety of search engines, and, after retrieving various search results, they finally determined to select the data from the Liao Liao Ting Web site as the dataset in the field of ceramics.

Development Environment

In this project, the authors used the Python language as the platform to realize the crawling of relevant data in the ceramic field. Table 1 shows the development environment.

Data Acquisition

The system mainly acquires data from the three columns of ceramic craftsmanship, ceramic art, and ceramic history. The system crawler uses Request, Beautiful Soup, and Pandas as platforms. After completing data acquisition, the authors used Pandas to save the data in CSV format. Figure 2 shows the schematic diagram of the data acquisition process.

Figure 1. Knowledge Graph and Corpus Building Process

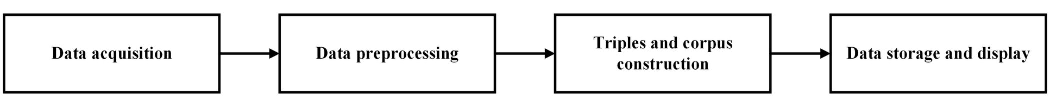
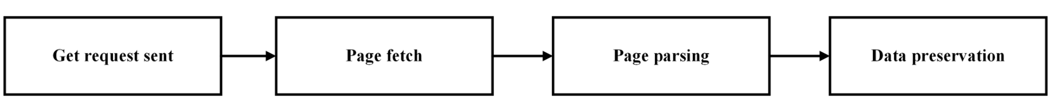


Table 1. Development environment

Operation system	Windows 11 x64
IDE	Visual Studio Code
Development language	Python 3.6.5
Browser	Chrome

Figure 2. Schematic diagram of the data acquisition process



Data Processing

When crawling the data, the authors did not deal with the redundant information in the data information in time, so the data contain a large number of illegal characters, which brings much disturbance to the subsequent model training and prediction, and increases the uncertainty factors of the model. In this section, the researchers mainly used Python language to remove this redundant and illegal interference, so as not to cause trouble to the downstream related algorithm models due to the disordered data source. Table 2 is an example of illegal character processing.

Workflow of Knowledge Graph Application in the Ceramic Field

The workflow of natural language processing-based knowledge graph application in the ceramic field encompasses two main tasks: Entity recognition and relationship recognition. In this section, the authors will focus on two algorithm models, namely BiLSTM-CRF and TextCNN, which are used to implement entity recognition and relationship recognition, respectively.

When users enter a question, they must first segment the word and then match the dictionary entity. Then, the two models are processed, and, finally, the system queries the answer required by the user, according to the result of the model processing. Figure 3 shows the specific process.

BiLSTM-CRF Named Entity Recognition Algorithm

- **BiLSTM Module:** The long short-term memory (LSTM) network can handle the gradient explosion phenomenon that occurs in recurrent neural networks. Three structures are added to LSTM: Memory gate, forget gate, and output gate; the memory gate determines whether the information is stored, the forget gate determines whether the information is forgotten, and the output gate is used to judge the current state.

The LSTM model is a one-way structure, and cannot encode information from right to left or from back to front in the process of sentence modeling, so it is not ideal for contextual and semantic information processing. BiLSTM is a bidirectional long-term and short-term memory network, which can better capture bidirectional dependencies, and uses forward LSTM and backward LSTM to process each word sequence, so that each moment feature has forward and backward dependencies. Figure 4 shows the BiLSTM network structure.

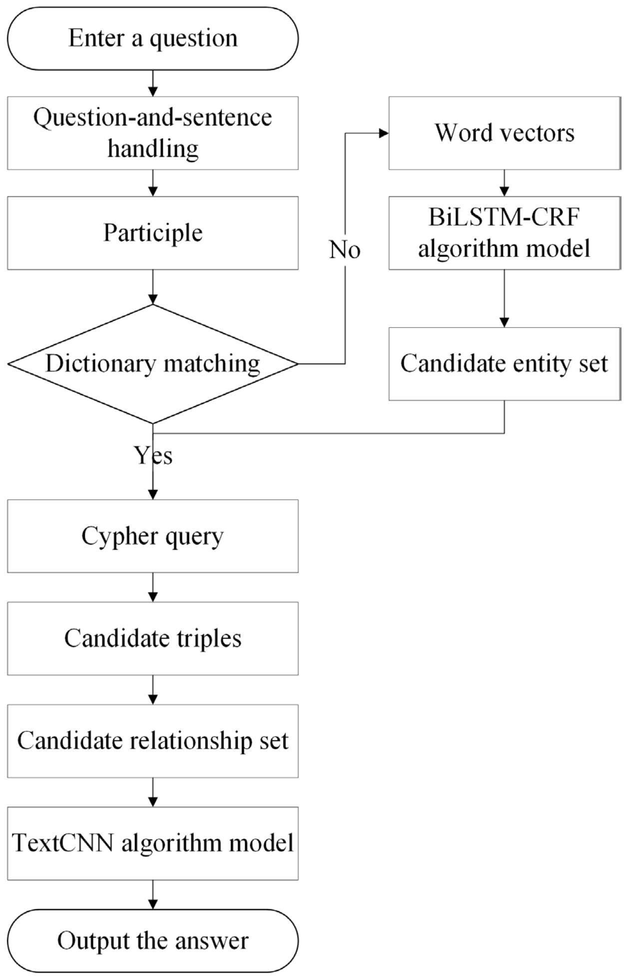
- **CRF Conditional Random Field Module:** A conditional random field (CRF) can obtain the best prediction sequence through the relationship between adjacent labels. As for the BiLSTM-CRF algorithm model, the function of the CRF is to predict the output sequence through BiLSTM to optimize the objective function.

For the input sequence $X=(X_1, X_2 \ldots X_n)$, the predicted output sequence is $Y=(Y_1, Y_2 \ldots Y_n)$. The score can be expressed as Equation 1, that is, the transition probability is added to the state probability:

Table 2. Example of illegal character handle

Original characters	Processing
Spaces	Auto
Tabs	Auto
Chinese characters	Manual
Garbled character	Auto

Figure 3. Workflow diagram for the application of knowledge graphs in the field of ceramics



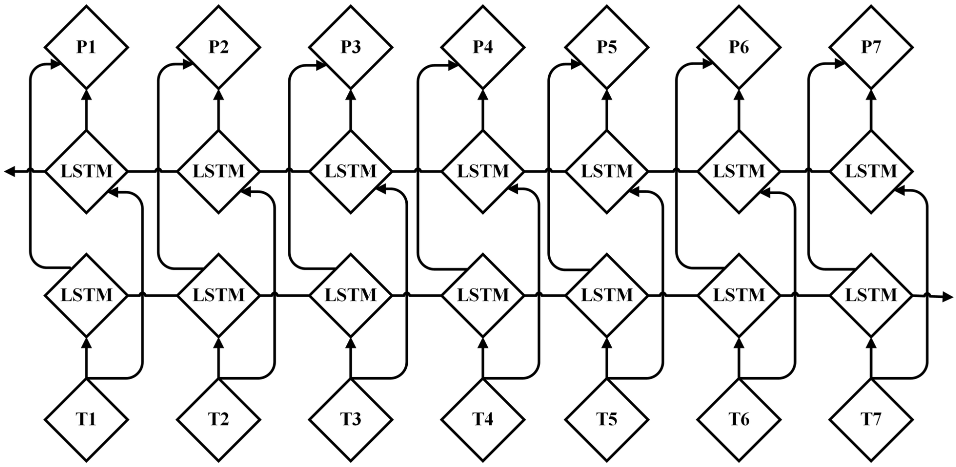
$$S\left(X,y\right)=\sum\nolimits_{i=0}^{n}A_{yi,yi+1}+\sum\nolimits_{i=1}^{n}P_{i,yi} \tag{1}$$

A is used to represent the transition matrix, and P is used to represent the output score matrix of BiLSTM. The probability value of the label sequence Y is obtained by SoftMax:

$$p(y\mid X)=\frac{e^{S\left(X,y\right)}}{\sum\nolimits_{y'\in Y}X^{e^{S\left(X,y'\right)}}} \tag{2}$$

Each node of the CRF network represents the predicted value, respectively. According to the predicted sequence output by BiLSTM, the path with the highest probability in the network is searched, and the output named entity is labeled and identified to complete the named entity identification. Therefore, the training aims at maximizing the probability P(y|X). This training aims at maximizing the probability, which can be achieved by the log-likelihood as shown below:

Figure 4. Structure diagram of the BiLSTM network



$$\log\left(p(y \mid X) = \log \frac{e^{S(X,y)}}{\sum_{y'\in Y} X^{e^{S(X,y')}}} = S\left(X,y\right) - \log\left(\sum_{y'\in Y} X^{e^{S(X,y')}}\right)\right) \tag{3}$$

The optimal path of the solution is obtained by predictive decoding through the Viterbi algorithm:

$$y^* = \arg\max score\left(x, y'\right) \tag{4}$$

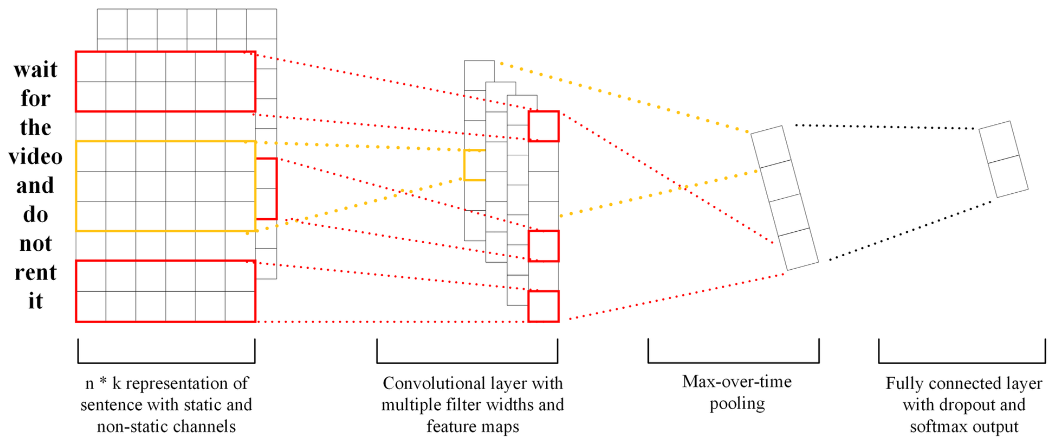
TextCNN Relation Recognition Algorithm

After the named entity recognition is completed, the relationship recognition is a matching and recognition operation between the relationship attributes associated with each entity and the corresponding relationship attributes on the knowledge graph. For example, the researcher asks a question: “What is the alias of blue and white porcelain?” After determining the entity label “blue and white porcelain,” they find the related attributes of the entity from the knowledge graph, such as “introduction,” and “alias”. In this study, the authors applied the TextCNN algorithm model to relation recognition. Unlike recurrent neural network and other sequence models, TextCNN has a simple network structure, but the introduction of trained word vectors can still achieve very good results and has a very fast training speed. First, the sentence context features are extracted, then they are sent to the TextCNN network for convolution operation to obtain the semantic vector of both the question sequence and the candidate relationship attribute, and, after this, the similarity calculation can be used to obtain the relationship recognition result. The structure of TextCNN is relatively simple and the training volume is faster. Figure 5 shows the network structure of TextCNN.

The input to the above image is the embedding layer obtained using the pretrained word vector (Word2Vector or glove) method. Each word vector is trained in an unsupervised manner.

In the past, CNN was generally regarded as a work in the Computer Vision(CV) field and applied in the direction of computer vision, but Yoon Kim partially deformed the CNN input layer and proposed a text classification model TextCNN. Compared with the traditional image CNN, the network structure of

Figure 5. Structure diagram of the TextCNN network

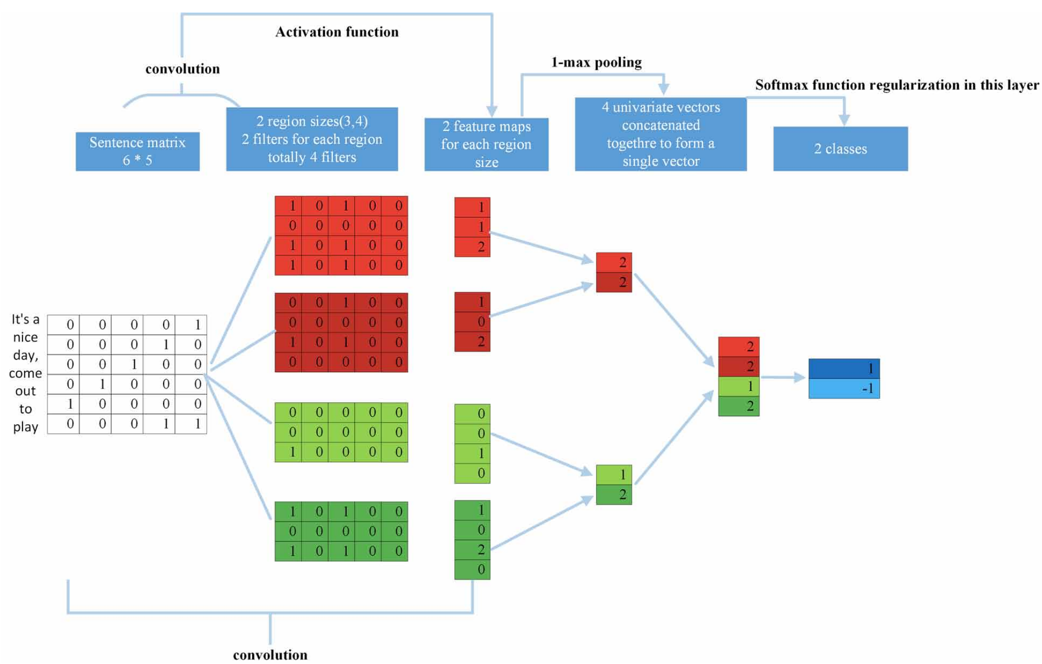


the TextCNN has not changed much. As Figure 6 shows that TextCNN actually only needs one layer of convolution and one layer of max-pooling, and its final output is connected to SoftMax for n classifications.

After the above analysis, the authors used the cosine similarity to calculate the similarity of the semantic vector between the question and the relation attribute. Equation 5 shows the calculation method:

$$sim\frac{\sum_{i=1}^{n} x_i y_j}{\sqrt{\sum_{i=1}^{n} x_i^2}\sqrt{\sum_{i=1}^{n} y_i^2}}_{\cos} \tag{5}$$

Figure 6. TextCNN schematic



In Equation 5, x_i and y_i are the i element of the semantic vector of the question sentence, and the j element of the semantic vector of the attribute candidate set. The reason why the cosine similarity is selected for calculation is that it is mainly affected by the direction. The smaller the included angle, the smaller the cosine distance and the vector length does not interfere much with its calculation.

EXPERIMENTAL AND EVALUATION

Experimental Environment and Evaluation Indicators

In this research, the authors used TensorFlow to build the experimental model. Table 3 shows the configuration of the experimental hardware and software environment.

The authors used the precision rate, recall rate, and F1 value as indicators to judge the accuracy of the model. The calculation formula is as follows:

$$P = \frac{TP}{TP + FP} \times 100\% \tag{6}$$

$$R = \frac{TP}{TP + FN} \times 100\% \tag{7}$$

$$F1 = \frac{2PR}{P + R} \times 100\% \tag{8}$$

Among them, TP represents the number of correctly identified samples; FP represents the number of wrongly recognized samples, FN represents the number of samples that were not identified, P is the precision rate, and R is the recall rate.

In entity recognition, the commonly used annotation modes are BIO mode, BIOE mode, and BIOES mode. In the experiment in this study, the authors adopted the BIO mode, which has three tags: “O,” “B,” and “I,” where O is a nonnamed entity, B is the first word of a named entity, and I is a nonfirst word of a named entity.

Entity Recognition Experiment

This experiment is mainly divided into two parts. To prove the generality of the model in this research, the authors verified it on the Own Think dataset and the ceramic field dataset they had previously constructed. Table 4 shows the pseudocode of the construction process of the ceramic field dataset.

The Own Think dataset is issued by the Thinking team and is the largest 140 million Chinese knowledge graph dataset publicly available in China. The Own Think dataset contains a total of seven kinds of labels in three categories: Locations name, Organization name, and Person name. The training and test sets of both datasets are split 10:1.

According to Figure 7, different experimental parameters have a greater impact on the F1 value. Therefore, according to the test results in Figure 7, the experimental parameter with the largest F1 value is selected as the parameter setting of the experiment (Tables 5 and 6).

Table 3. Experimental environment configuration

Memory	DDR4 20G
GPU	RTX 3060
Development language	Python 3.6.5
TensorFlow	1.14.0

Table 4. Algorithm for constructing an annotated set of entity recognition data in the ceramic field

Input: Q&A training set, knowledge graph. Output: Question entity annotation set. Process: 1. Find triple data in the knowledge graph according to the standard answer of the question. 2. Determine the relevant entity name in the question according to the found triple. 3. Mark "1" for the word sequence that coincides with the entity name in the question sequence, and mark "0" for the rest, and output the marked question sequence for saving.

Figure 7. The effect of different parameter settings on the F1 value in two datasets

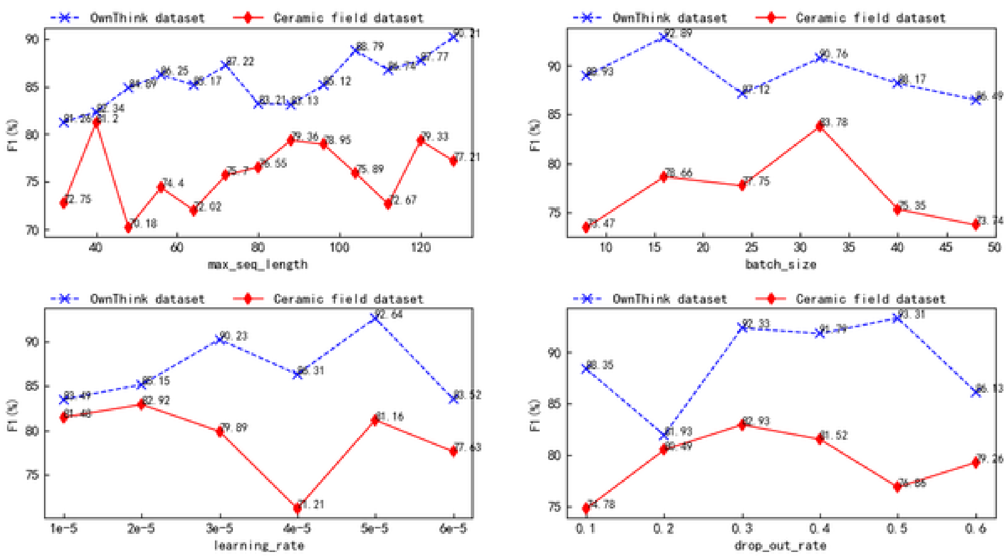


Table 5. Experimental parameters of the own think dataset

Parameters	Value
max_seq_1ength	128
batch_size	16
learning_rate	5e-5
drop_out_rate	0.5

Table 6. Experimental parameters of the ceramic field dataset

Parameters	Value
max_seq_1ength	40
batch_size	32
learning_rate	2e-5
drop_out_rate	0.3

In this study, the authors used the BiLSTM-CRF algorithm model for entity recognition operation. To prove the effectiveness of the model, they compared it with models such as LSTM-CRF. Table 7, Table 8, and Figure 8 show the specific experimental results.

As Table 7, Table 8, and Figure 8 above show, the BiLSTM-CRF algorithm model has achieved good results on both the ceramic field dataset and the OwnThink dataset, which proves that the model has certain versatility and effectiveness; especially in this study, the BiLSTM-CRF algorithm

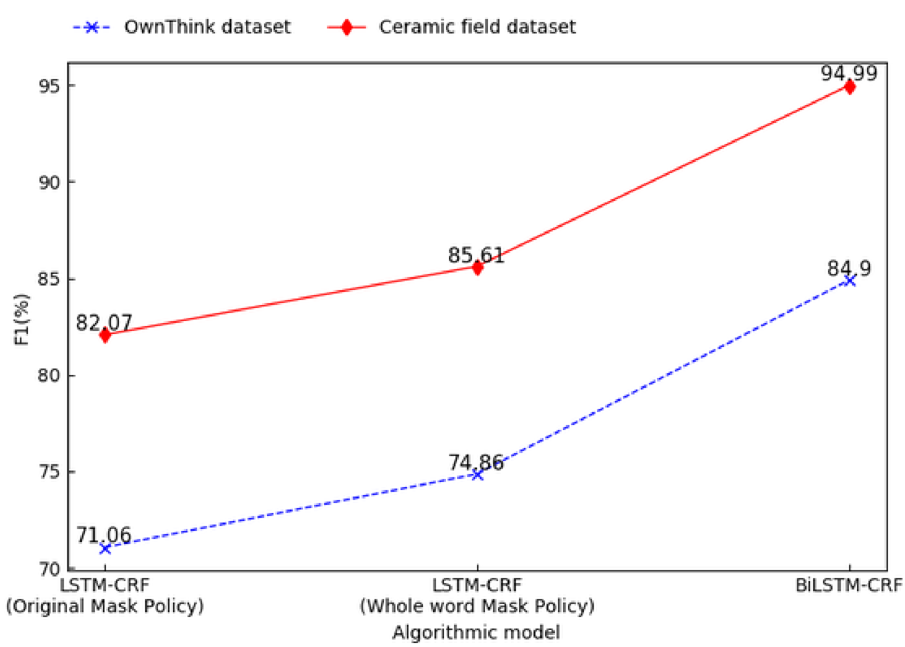
Table 7. Experimental results of the ceramic field dataset

Model	Evaluation Indicators		
	Precision	Recall	F1
LSTM-CRF (Original Mask Policy)	72.43	69.75	71.06
LSTM-CRF (Whole word Mask Policy)	75.37	74.36	74.86
BiLSTM-CRF	84.78	85.03	84.90

Table 8. Experiment results of the own think dataset

Model	Evaluation Indicators		
	Precision	Recall	F1
LSTM-CRF (Original Mask Policy)	83.44	80.75	82.07
LSTM-CRF (Whole word Mask Policy)	86.75	84.50	85.61
BiLSTM-CRF	94.89	95.10	94.99

Figure 8. Experimental results of the ceramic field dataset and the own think dataset



model can significantly improve the effect of entity recognition. At the same time, the application of the LSTM-CRF model based on the whole-word MASK strategy and the document-level attention mechanism has improved the performance of the model to varying degrees. In the ceramic field dataset, all models have to go through many iterations in the initial stage to obtain a relatively stable level, which is also related to the excessive number of specific entity names in the ceramic field.

This experiment also allowed to compare the performance of the model the authors used in this study with other existing deep learning models on the OwnThink and ceramic field datasets (Tables 9 and 10).

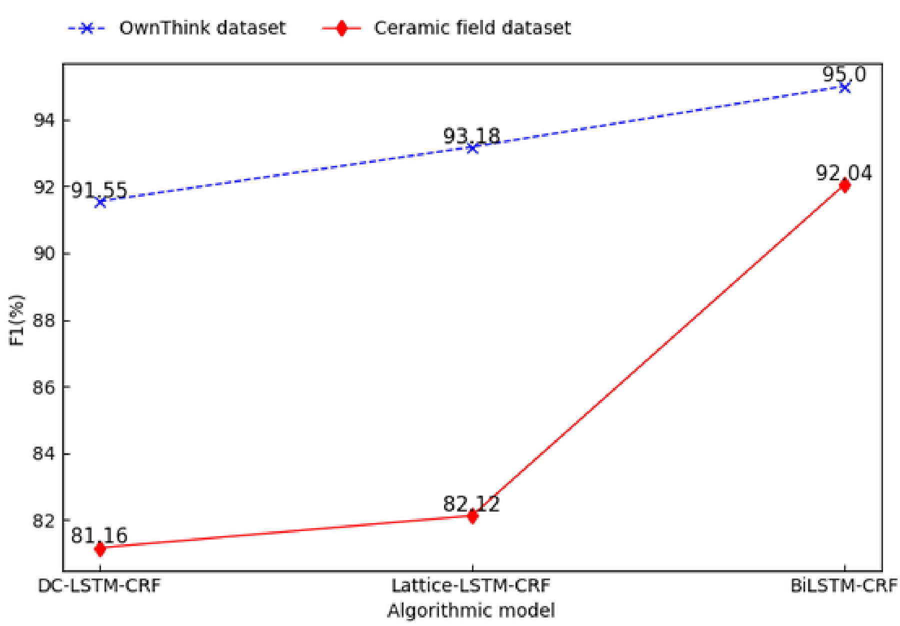
Table 9. Comparison of existing models on the own think dataset

Model	Result		
	Precision	Recall	F1
DC-LSTM-CRF (Liu et al., 2020)	92.14	90.96	91.55
Lattice-LSTM-CRF (Zhang & Yang, 2019)	93.57	92.79	93.18
BiLSTM-CRF	94.89	95.10	95.00

Table 10. Comparison of existing models on the ceramic domain dataset

Model	Result		
	Precision	Recall	F1
DC-LSTM-CRF (Liu et al., 2020)	81.46	80.87	81.16
Lattice-LSTM-CRF (Zhang & Yang, 2019)	83.17	81.09	82.12
BiLSTM-CRF	92.73	91.35	92.04

Figure 9. Experimental results of the ceramic field dataset and the own think dataset



As Table 9, Table 10, and Figure 9 above show, compared with other existing models, the model of this study achieves the best results in accuracy, recall, and F1 value on the Own Think dataset, and the accuracy, recall, and F1 value on the ceramic field dataset. The rate and F1 value are also significant. It is evident that the BiLSTM-CRF model is more versatile and effective.

Relation Recognition Experiment

The authors used the TextCNN algorithm for relationship recognition. First, they used the Text model as the word embedding to extract the contextual features of the sentence. Differently from the entity recognition model, the researchers input the feature information output by Text into the CNN, do the convolution operation on the question sentence and the relationship vector respectively, and finally, calculate the cosine similarity of the semantic vectors of both of them, which is used to complete the recognition of the entity relationship.

According to Figure 10, different experimental parameters have a greater impact on the F1 value. Therefore, according to the test results in Figure 10, the authors selected the experimental parameter with the largest F1 value as the experimental parameter setting (Table 11).

The dataset the authors used in this experiment is the self-built ceramic field dataset of this study. Table 12 presents the algorithm pseudocode constructed by the ceramic field relationship identification data label set.

Figure 10. The Influence of different parameter settings on the f1 value in the ceramic domain dataset

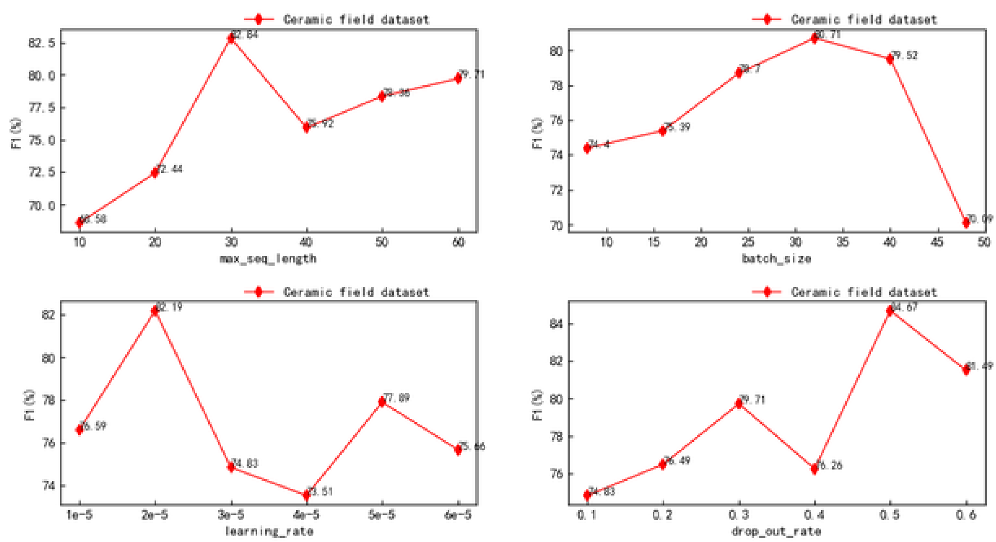


Table 11. Experimental parameters of relationship recognition

Parameters	Value
max_seq_1ength	30
batch_size	32
learning_rate	2e-5
drop_out_rate	0.5

Table 13 shows the experimental results.

Table 13 and Figure 11 above show that the model in this study has achieved better results than the original Text model without the whole-word MASK strategy. This further proves that, for Chinese data, it is feasible and effective to adopt the whole-word MASK strategy adapted to Chinese. At the same time, the model also achieved the best results in this experiment, with F1 reaching 83.90%. On the ceramic

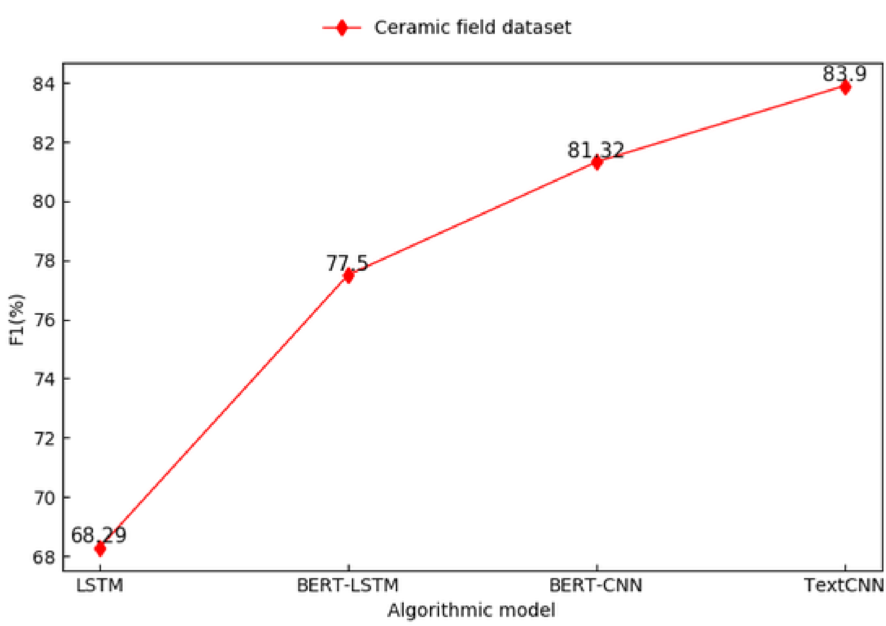
Table 12. Algorithm for constructing an annotated set of data for relationship recognition in the ceramic field

Input: Q&A training set, knowledge graph. Output: Question entity data annotation set. Process: 1. Find triple data in the knowledge graph according to the standard answer of the question. 2. Determine the relevant relationship attribute of the question according to the found triple data and the entity name determined in the entity recognition. 3. Label the relational attributes in the question sequence.

Table 13. Experimental results of relationship recognition

Model	Result		
	Precision	Recall	F1
LSTM	72.51	64.53	68.29
BERT-LSTM	78.59	76.44	77.50
BERT-CNN	82.55	80.12	81.32
TextCNN	84.51	83.29	83.90

Figure 11. Experimental results of relationship recognition of different algorithm models in the dataset of ceramics



field dataset, the model is not only stable, but also has the highest accuracy, compared with other models. However, all models, including the model in this paper, can only get a relatively stable level after many iterations, which is related to the relative complexity of the relationship in the ceramic field. The performance of the authors' algorithm model in the ceramic field dataset also has much room for improvement.

Evaluation

This section mainly introduces two algorithm models, namely the entity recognition algorithm BiLSTM-CRF and the related recognition algorithm TextCNN. In entity recognition, the authors used a combination of algorithm and dictionary matching and validated its effectiveness on two datasets. They also compared the BiLSTM-CRF with other existing models. After comparing the two datasets, the BiLSTM-CRF algorithm performs relatively well in the general field, but not in the ceramic field. The research following this study will provide directions to adapt the algorithmic model to the characteristics of the ceramic field. In relation to recognition, in this paper the author proposed the TextCNNT algorithm model. The difference is that the features extracted from Text are input into the CNN for convolution operation to extract corpus features. The experimental results show that the generality and effectiveness of the authors' algorithm model have achieved good experimental results.

APPLICATION OF KNOWLEDGE GRAPH IN THE CERAMIC FIELD

Architecture Design of Ceramic Knowledge Question Answering System

Figure 12 shows the architecture diagram of the ceramic knowledge question and answer system.

Ceramic Knowledge Question Answering System

The authors built a simple system page and completed the questioning operation through the input box in the dialog interface. After the system finds the answer, it will directly return the answer and display it on the page. Due to the connection with the speech synthesis module, the system will present the answer in both text and voice (Figure 13).

CONCLUSION

In this paper, the authors preferred to obtain relevant unstructured data in the ceramics field through crawler technology in the Internet. They constructed and validated an algorithm model suitable for nomenclature recognition and relationship recognition in the ceramics field through experiments,

Figure 12. Architecture design of the ceramic knowledge quiz system

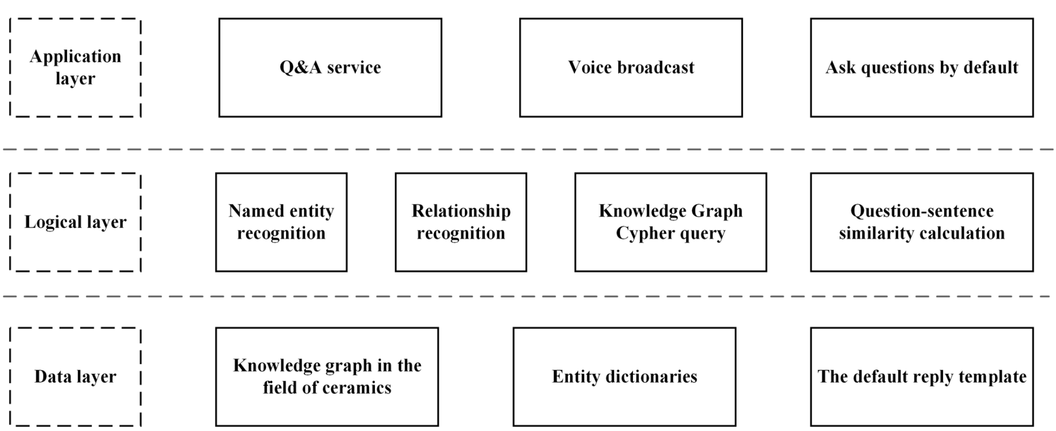
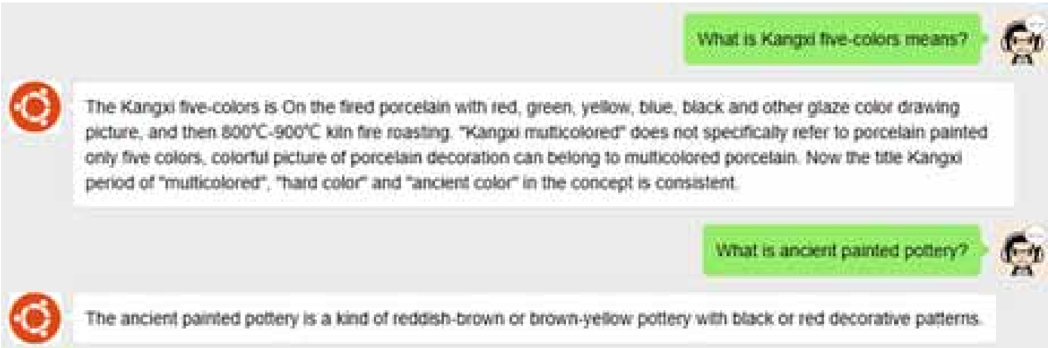


Figure 13. Example of a ceramic knowledge question-and-answer system



and implemented a Q&A system based on the knowledge graph in the ceramics field. In brief, the authors' work was as follows:

1. They carried out the acquisition of relevant data and the construction of datasets according to the needs of ceramics domain workers.
2. They conducted the relationship recognition and attribute recognition experiments by combining the OwnThink dataset and the self-constructed ceramic field dataset, and then selected the relevant algorithm models for the construction of the ceramic field knowledge graph.
3. By using the constructed ceramic field knowledge graph, they completed the development of an intelligent Q&A service for ceramic knowledge.

The researchers made a bold attempt in the construction and application of the knowledge graph in ceramics field. This paper offers the following innovative points:

1. Scale-up acquisition of unstructured data in the ceramics field. The authors used the Beautiful Soup framework in Python language to crawl the ceramic data of relevant sites in a large scale, and obtained the ideal plain text data after data preprocessing.
2. Semiautomatic construction of a knowledge graph in the ceramics domain. The authors used deep learning to extract association triples and construct the knowledge graph semiautomatically, which saves the cost of knowledge graph construction, improves the efficiency of knowledge graph, and fills the gap of knowledge graph in ceramics field.
3. The developed ceramic knowledge intelligent question-and-answer service can accomplish more accurate ceramic knowledge data retrieval.

Finally, through this research, the authors found that most of the commonly used entity recognition and relationship recognition algorithm models require high computing power of equipment and large computational workload. In their future work, they will further optimize and improve the rate and quality of knowledge graph construction in ceramics field.

AUTHOR CONTRIBUTIONS

YN carried out most of the experiments and of the article writing. YZ and NH were responsible for part of the majority of the idea, as well as part of the writing and providing computational resources. GS and JP partially provided the idea, conducted some of the experiments (e.g., data processing and analysis experiments), and offered important suggestions. YP and CN conducted part of the experiments and performed proofreading.

FUNDING

This work was supported by the Jingdezhen City Social Science Planning Project (No.202233), the Science and Technology Research Project of Jiangxi Provincial Department of Education (No. GJJ2201032, GJJ2201049, GJJ2204703, GJJ211348), and the Science and Technology Project of Jingdezhen (No. 20212GYZD009-02); Chaozhou Ceramic Industry Talent Revitalization Plan (No. 2021YJ03).

DATA AVAILABILITY

The authors will make all the data and material publicly available upon the acceptance of their work.

CODE AVAILABILITY

The authors will make all the codes publicly available upon the acceptance of their work.

CONFLICT OF INTEREST

The authors declare that there are no conflicts of interest regarding the publication of this paper.

CONSENT TO PARTICIPATE

All authors have agreed the consent to participate in the work.

CONSENT FOR PUBLICATION

All authors have checked the manuscript and have agreed to the publication.

REFERENCES

Cao, X. (2021). *Research and application of knowledge map of blue and white porcelain*. Beijing University of Posts and Telecommunications.

Fu, L., Cao, Y., Bai, Y., & Leng, J. (2021). Development status and prospect of vertical domain knowledge graph in China. *Computer Applied Research*, *38*(11), 3201–3214.

Guo, X., Ma, L., Su, K., & Sun, Y. (2018). Research on automatic metadata quality evaluation method of data repository based on semantic annotation. *Computer Applications and Software, 35*(6), 23—27+82.

Hang, P. X., Zhao, X., Fang, Y., Zhu, H., Xiao, W., & College, T. (2019). End-to-end knowledge triplet extraction combined with adversarial training. *Journal of Computer Research and Development*, *56*(12), 2536–2548.

Katiyar, A., & Cardie, C. (2017). Going out on a limb: Joint extraction of entity motions and relations without dependency trees. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*. (vol. 1, pp. 917—928). Association for Computational Linguistics.

Kwiatkowksi, T., Zettlemoyer, S., & Goldwater, M., Steedman. (2010). Inducing probabilistic CCG grammars from logical form with higher-order unification. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (*vol. 1, pp. 1223–1233). Association for Computational Linguistics.

Li, Q., & Ji, H. (2014). Incremental joint extraction of entity mentions and relations. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics* (vol. 1, pp. 402—412). Association for Computational Linguistics. doi:10.3115/v1/P14-1038

Li, X. (2020). Countermeasures for the construction of ceramic cultural resources database. *Digital World*, *182*(12), 83–84.

Liang, P., Jordan, M. I., & Klein, D. (2011). Learning dependency-based compositional semantics. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies* (vol. 1, pp. 590—599). Association for Computational Linguistics.

Liu, L., Shi, W., & Cheng, S. (2016). A review of the application of knowledge graph in question answering system. *Automation technology and applications, 39*(06), 173—176.

Liu, X., Gu, L., & Shi, X. (2020). Named entity recognition based on Bi-LSTM and attention mechanism. *Journal of Luoyang Institute of Technology (Natural Science Edition), 29*(1), 65—70,77.

Ruan, T., Wang, M., Wang, H., & Hu, F. (2016). Construction and application of vertical knowledge graph. *Knowledge Management Forum, 1*(3), 226—234.

Wang, C., Yu, H., & Wan, F. (2018). Information retrieval technology based on knowledge graph. In *Proceedings of the 2018 3rd International Conference on Advances in Materials, Mechatronics, and Civil Engineering (ICAMMCE 2018)* (vol. 162, pp. 300—305). Atlantis Press. doi:10.2991/icammce-18.2018.65

Wang, Y., Chen, G., Cai, D., & Wang, P. (2016). Construction of term semantic knowledge base based on HowNet. *Journal of Shenyang Aerospace University*, *33*(4), 78–84.

Weizenbaum, J. (1966). ELIZA: A computer program for the study of natural language communication between man and machine. *Communications of the ACM*, *9*(1), 36–45. doi:10.1145/365153.365168

Wu, X., & Huang, Y. (2021). Construction and promotion of ceramic culture inheritance platform under the background of big data. *Digital Technology and Applications*, *39*(11), 193–195.

Xia, Y. (2018). *Research on agricultural knowledge service system based on knowledge graph*. Anhui Agricultural University.

Xu, Z., Sheng, Y., He, L., & Wang, Y. (2016). A review of knowledge graph technology. *Journal of University of Electronic Science and Technology of China*, *45*(4), 589–606.

Yao, X., & Durme, B. V. (2014). Information Extraction over Structured Data: Question Answering with Freebase. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics* (vol. 1, pp. 956–966). Association for Computational Linguistics. doi:10.3115/v1/P14-1090

Zeng, X. R., Zeng, D. J., He, S. Z., Liu, K., & Zhao, J. (2018). Extracting relational facts by an end-to-end neural model with copy mechanism. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics* (vol. 1, pp. 506—514). Association for Computational Linguistics. doi:10.18653/v1/P18-1047

Zettlemoyer, L. S., & Collins, M. (2012). Learning to map sentences to logical form: Structured classification with probabilistic categorial grammars. In *Proceedings of the Twenty-First Conference on Uncertainty in Artificial Intelligence* (vol. 1, pp. 658–666). AUAI Press.

Zhang, K., Li, W., & Wang, H. (2015). Ontology-based question answering system in aviation field. *Chinese Journal of Information*, *29*(4), 192–198.

Zhang, Y., & Yang, J. (2019). Chinese NER using lattice LSTM. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics* (vol. 1, pp. 1554–156). Association for Computational Linguistics.

Zheng, S. C., Hao, Y. X., Lu, D. Y., Bao, H., Xu, J., Hao, H., & Xu, B. (2017). Joint entity and relation extraction based on a hybrid neural network. *Neurocomputing*, *257*, 59–66. doi:10.1016/j.neucom.2016.12.075

Zheng, S. C., Wang, F., & Bao, H. Y. (2017). Joint extraction of entities and relations based on a novel tagging scheme. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics* (vol.1, pp. 1227—1236). Association for Computational Linguistics. doi:10.18653/v1/P17-1113

Zhu, J., Yang, Y., Xu, B., & Li, J. (2019). Semantic representation learning based on HowNet. *Journal of Chinese Information Science*, *33*(3), 33–41.

Yu Nie, a member of Jiu San Society, lecturer/software developer (full stack), graduated from Jingdezhen Ceramic Institute with a bachelor's degree in computer science and technology in 2007, and graduated from Nanchang University with a master's degree in computer technology in 2013. He joined Jingdezhen Ceramic University in 2014, was a visiting scholar at Wuhan University in 2020, and entered Wuhan University for his Ph.D. He mainly researches on privacy-preserving computing, natural language processing, and intelligent applications related to machine vision. He has completed more than 40 projects and published more than 40 academic papers, including 12 SCI/EI papers and 2 core Chinese journals. He has obtained 3 authorized patents and 40 software copyright registrations.

Na Huang, teaching assistant and e-commerce teacher, graduated from Nanchang University College of Science and Technology, majoring in e-commerce with a bachelor's degree in management in 2015, and joined Jingdezhen Vocational University of Arts in 2022, teaching e-commerce. In her daily teaching and research work, her rigorous teaching attitude and sincere treatment of students have been widely praised and loved by colleagues and students. She mainly researches natural language processing and intelligent recommendation model research. At present, she has completed three projects of science and technology projects for the Jiangxi Education Department, one project of science and technology for Jingdezhen City, and published 2 SCI/EI indexed papers.

Junjie Peng, an undergraduate, entered the School of Information Engineering, Jingdezhen Ceramic University, in 2020, majoring in computer science and technology, and was granted a scholarship six times. During his school years, he works as a researcher at the Jiangxi Province Engineering and technology research center of Ceramic Enterprise Informatization. He mainly followed Mr.Nie to engage in natural language processing research and computer software application development, participated in 5 projects of the College Students' Innovative Entrepreneurial Training Plan Program, and won several discipline competitions awards. He mainly studies on natural language processing and intelligent applications and will continue to study artificial intelligence after graduation from Jingdezhen Ceramic University.

Guanghua Song entered the School of Information Engineering, Jingdezhen Ceramic University, in 2018, majoring in computer science and technology, and graduated from Jingdezhen Ceramic University in 2022 with a bachelor's degree in engineering in computer science and technology. During his school years, he mainly followed Mr.Nie to engage in natural language processing research and computer software application development, worked as the president of the Mobile Internet Association, completed three projects of the Student Innovation Program, received one national scholarship, won several discipline competitions awards, wrote his graduation thesis mainly about the knowledge quiz in the field of ceramics, and currently works mainly on natural language processing.

Yilai Zhang received the B.S. degree in silicate engineering from Jingdezhen Ceramic Institute, Jingdezhen, China, in 1986 and the PhD in calculation material from South China University of Technology, Guangzhou, China, in 2008 .Professor II of Jiangxi Province, Director of Jiangxi Ceramic Enterprise Informatization Engineering Technology Research Center of Jingdezhen Ceramic University, Head of Postgraduate Teaching Supervision Group of Jingdezhen Ceramic University, Member of the First Discipline Review Group of Degree Committee of Jiangxi Provincial People's Government, Expert of Master's Professional Degree Expert Committee of Degree Committee of Jiangxi Provincial People's Government, Deputy Head of Manufacturing Informatization Expert Group of Jiangxi Province, Innovation and Entrepreneurship of Jiangxi University Students In 2001, he was an advanced individual of national teacher morality, a national outstanding teacher in 2014, an advanced worker of Jiangxi Province in 2005 (provincial model worker), and a special allowance of the provincial government.

Yongkang Peng, professor, Master tutor. Young and middle-aged backbone teacher of Jiangxi Provincial colleges and universities, member of the Mathematics Teaching Steering Committee of Jiangxi Provincial Colleges and Universities from 2021 to 2024. He received his Bachelor's degree in Mechanical Design and Manufacturing from Jingdezhen Ceramic Institute in 1993 and his Master's degree in Computer Application Technology from Nanchang University in 2008. His research focuses on computer application software technology, CAD/CAM technology in the ceramic industry, artificial intelligence, and computer vision. He has completed more than 20 projects, published more than 20 academic papers, three software Copyrights, and two patents.

Chenglin Ni is an engineer. He graduated from Jingdezhen Ceramic Institute, College of Science and Technology and Art with a Bachelor's degree in Inorganic Materials and Nonmetal Engineering in 2021 and graduated from Kunming University of Technology with a Master's degree in Materials Science and Engineering in 2014. He used to work in Haier Group Blue Whale Technology Co. Ltd. as deputy general manager, mainly engaged in ceramic production management and scientific and technological project research. He was awarded the second prize for innovation in China's building and sanitary ceramics industry, the second prize for science and technology invention by the Guangdong Light Industry Federation, and the special prize for technical innovation in the Guangdong building materials industry.