



Construction of Knowledge Graph Based on Discipline Inspection and Supervision

Yuefeng Liu, Wei Guo, Hanyu Zhang, Haodong Bian,
Yingjie He, Xiaoyan Zhang and Yanzhang Gong

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

May 26, 2021

Construction of Knowledge Graph Based on Discipline Inspection and Supervision

Yuefeng Liu

*School of Information Engineering
Inner Mongolia University of Science
and Technology*
Baotou, Inner Mongolia, China
liuyuefeng01035@163.com

Wei Guo

*School of Information Engineering
Inner Mongolia University of Science
and Technology*
Baotou, Inner Mongolia, China
940600846@qq.com

Hanyu Zhang

*School of Information Engineering
Inner Mongolia University of Science
and Technology*
Baotou, Inner Mongolia, China
2020942@imust.edu.cn

Haodong Bian

*School of Information
Engineering
Inner Mongolia University
of Science and Technology*
Baotou, Inner Mongolia,
China
609354432@qq.com

Yingjie He

*School of Information
Engineering
Inner Mongolia University
of Science and Technology*
Baotou, Inner Mongolia,
China
1353517169@qq.com

Xiaoyan Zhang

*School of Information
Engineering
Inner Mongolia University
of Science and Technology*
Baotou, Inner Mongolia,
China
1213185642@qq.com

Yanzhang Gong

*Inner Mongolia Discipline
Inspection and Supervision
Big Data Laboratory*
Baotou, Inner Mongolia,
China
1435283338@qq.com

Abstract—To solve the problems of large number of notifications, low relevance and no relevant knowledge base in the field of discipline inspection, a method of constructing a knowledge map of discipline inspection and supervision based on the BERT-BiLSTM-CRF model is proposed. Firstly, the unstructured data is collected from the content of the disciplinary inspection and supervision report. Through the bottom-up method the notification concept layer is constructed. By using deep learning models to extract entities. Then the entities and semantic relations are stored in the graph database Neo4j and displayed in the form of a knowledge graph. This method realizes the whole process from unstructured data to knowledge graph, and provides technical reference for the construction of domain-based knowledge graph. Simultaneously, the knowledge map in the discipline inspection field established through examples provides support and assistance for the discipline inspection personnel's scientific decision-making.

Keywords—discipline inspection and supervision, knowledge graph, *NER*(natural language processing), *BERT* (*Bidirectional Encoder Representations from Transformers*), graph database

I. INTRODUCTION

Discipline inspection and supervision are two functions performed by the party's discipline inspection agency and the government's supervision department. Strengthening the construction of discipline inspection and supervision is a necessary task for strengthening the party's capacity building. With the further strengthening of work intensity, traditional information storage methods become unable to meet the needs of staff in processing information. Constructing a knowledge map of discipline inspection and supervision is of great help to provide knowledge access and decision-making reference for practical anti-corruption works. The knowledge graph is a technique that adopts graph models to describe the knowledge and has a wide range of applications in assisted search, assisted question and answer, etc. By area of application, knowledge graphs can be categorized as general knowledge graphs and domain knowledge graphs. Among knowledge graphs, domain knowledge graphs have comparatively large knowledge sources, complex knowledge structures, and extensive

knowledge application forms, so studying its construction technology is of great significance. Reference [1] adopts core technologies including knowledge modeling, knowledge extraction, and knowledge fusion to construct a knowledge map in public security, and summarizes and integrates massive multi-source heterogeneous public security data. Reference [2] builds a multi-source data fusion military knowledge map based on actual application requirements. Reference [3] builds a knowledge map in the legal field through judgments, which helps legal practitioners to reason about cases more quickly and conveniently. However, no known reference applies knowledge graphs in discipline inspection, and in the discipline inspection field, there is no query-related database based on knowledge graphs. The introduction of knowledge graphs into disciplinary inspection is an innovative idea to answer how staff obtains key knowledge from a large amount of anti-corruption data.

The process of building a knowledge graph is generally divided into steps including knowledge modeling, knowledge extraction, and knowledge storage. In the process, the role of knowledge extraction is to obtain the association between entities. At present, common knowledge extraction methods include machine learning and deep learning. Deep learning is an emerging method in machine learning. It has powerful automatic feature extraction capabilities and advantages in processing high-dimensional and non-linear data. Therefore, in this study, research on knowledge extraction methods based on deep learning was carried out. After extracting the entities, the data was stored in the graph database to construct a knowledge graph in discipline inspection, and provide knowledge access for strengthening the party's capability of governing and promoting anti-corruption work.

II. EASE OF USE "DISCIPLINE INSPECTION AND SUPERVISION" KNOWLEDGE GRAPH CONSTRUCTION TECHNOLOGY

A. Dataset Establishment

Dataset acquisition is the first step in constructing a knowledge graph. Reference [4] adopts a crawler

framework to collect structured and unstructured data from websites including Global Military Network to complete and supplement the dataset. Reference [5] introduces the essential facts and civil adjudication theory, to build a legal knowledge map for knowledge accumulation. Data in disciplinary inspection comes from government administrative documents, magazines, newspapers, headlines, etc. In this study, the column on mental issues in the supervision and exposure section of the State Supervision Commission of the Central Commission for Discipline Inspection was selected as the data source. This data source contains thousands of disciplinary and illegal texts, covering the area, the person, the occurrence, disciplinary measures, etc., of the problems, which generally meets the entity content and entity attributes required for the construction of the knowledge graph.

B. Knowledge Modeling

Knowledge modeling design is similar to the table structure design of relational databases. It is the fundamental stage of the subsequent construction steps. The quality of knowledge modeling design is related to the success or failure of the entire knowledge graph construction. There are many design methods for knowledge graph modeling. Modeling should start from practical problems. For example, when a user queries the name of a department, it is necessary not only to understand which person under the department has been punished but also to understand which area the department with the same name belongs to; when inquiring about the problem, the corresponding disciplinary measures in the same situation and the information of the person who has had the problem should be displayed.

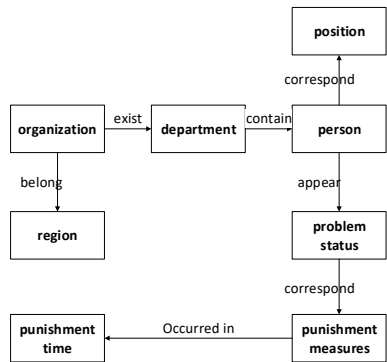


Fig. 1. “Discipline inspection and supervision” knowledge graph model

Reference [6] presents the central entity node and twelve first-level nodes through information comparison and analysis from websites including China Civilization Network and People’s Daily Online by constructing the knowledge map of socialist core values. Reference [7] formulates the ontology design scheme through the spiral feedback method based on demand in the military. Modeling can be done through eight core concepts of military weapons and common concepts including countries and production units. Reference [8] analyzes elements in public security explosion protection, presents the construction of the knowledge graph through the three central nodes including entities, events, and documents together with the determination of the respective categories and subcategories based on real cases and books in the respective curriculum. This study referred to the design ideas of knowledge modeling in other fields and summarized the existing and possible entities based on the content of the established dataset in Section A. For example, “Water Resources Bureau, community” is summarized as an organization, “disciplinary payment collection, the

expulsion of party membership” is summarized as punitive measures. After formulating the ontology design according to the entity, and performing the modeling design for the business of the discipline inspection and supervision knowledge map, the entity modeling diagram can be obtained as shown in Fig. 1.

C. Knowledge Extraction

By the column data and knowledge modeling, the entity of the discipline inspection and supervision database can be obtained, with contents including region, organization, department, person, position, problem status, punishment time, punishment measures, etc. In this study, using named entity recognition, the model was trained with labeled data, and the trained model was used to label unlabeled data. Although the format of the disciplinary inspection report data was relatively neat, there were many types of entities in the corpus. The method based on rules and dictionaries relies on domain experts to create rules and dictionaries. The extracted entities have a high accuracy rate, but they cannot cover the entire range and domains. It is difficult to adapt to new needs and new goals. The portability and accuracy of small sample data are not high. Therefore, in this study, deep learning was used to complete the named entity recognition task, and the BERT-BiLSTM-CRF model was constructed to complete the entity extraction. Bidirectional encoder representations from the transformers (BERT) [9] adopts transformer [10] as the main framework of the algorithm so that the transformer can more thoroughly capture the two-way relationship in the sentence, and the bidirectional long short-term memory (BiLSTM) can get bidirectional semantic information [11]. Conditional random field (CRF) can calculate the maximum probability sequence labeling [12]. In this study, through the artificially labeled corpus training model, the unlabeled corpus was labeled using the training model, and eight types of entities were separated from the notified corpus.

1) BERT: BERT stands for the two-way encoder representation from the transformer, which is characterized by fine-tuning through an additional output layer. It abandons the traditional recurrent neural network (RNN) and convolutional neural network (CNN) and converts the distance between two words at any position to 1 through the self-attention mechanism, thus effectively solving the problem of long-term dependence between characters. BERT completes the pre-training process through two self-supervised tasks. The first step is to randomly mask some words from the input corpus during training, and then predict the word through the context. The second step is to determine whether a sentence is the following of another sentence. If it is positive, it outputs “IsNext”, otherwise, outputs “NotNext”, so that the model understands the relationship between sentences. Through the above process, the accuracy of named entity recognition can be greatly improved. The model structure is shown in Fig. 2, where T stands for the transformer.

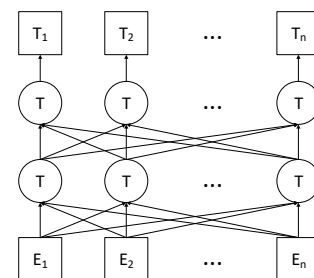


Fig. 2. Pre-trained language model by BERT

2) BiLSTM: Long short-term memory (LSTM) is a special type of RNN. Compared with RNN, LSTM can learn long-term dependent information by its “forget gate” and “memory unit” design features. For example, in the sentence “Zhang did not obtain funds through legal means”, “did not” is the negation of the following “legal means”, that is, Zhang’s behavior in this sentence is illegal. Using the LSTM model can help better capture the dependence of longer distances. The calculation is as follows.

$$f_t = \sigma(W_f h_{t-1} + W_f x_t + b_f) \quad (1)$$

$$i_t = \sigma(W_i h_{t-1} + W_i x_t + b_i) \quad (2)$$

$$o_t = \sigma(W_o h_{t-1} + W_o x_t + b_o) \quad (3)$$

$$C_t = i_t \tanh(W_c h_{t-1} + W_c x_t + b_c) + f_t C_{t-1} \quad (4)$$

$$h_t = o_t \tanh(C_t) \quad (5)$$

At time t , h_{t-1} and x_t represent the state of the hidden layer and the input of the current unit. In the formula, \tanh is the hyperbolic tangent function, σ is the activation function sigmoid, C_t is the state of the memory unit, i_t , f_t , and o_t correspond to the input gate, forget gate, and output gate of the LSTM, respectively. W and b represent the weight matrix and bias vector.

However, the use of LSTM to model sentences still has certain shortcomings, including the inability to encode back-to-front information, and the lack of reliance on the context in more fine-grained classification. BiLSTM solves this problem very well. The network adopts forward and backward LSTM output for each sequence to be spliced into a complete sequence, thus better obtaining the two-way semantic dependency.

3) CRF: The Bi-LSTM named entity recognition model can get the score prediction of the word corresponding to each category, but the prediction result lacks constraints. CRF can add some constraints to ensure the validity of the final prediction result, and these constraints are automatically learned during training data. For example, the beginning of a sentence should be “B-” or “O” instead of “I-”; the beginning of a person’s name should not be followed by the middle part of the organization name, so “B-PER I-PER” is correct, and “B-PER” “I-ORG” is wrong; “O I-DAT” is wrong because the starting position of the entity should be “B-” instead of “I-”. The loss function in the CRF layer includes two types of scores, emission scores (state scores) and transition scores. The emission score comes from the output of the BiLSTM layer and is the score of the prediction result. The transfer score obtains the score from one label to another by constructing a transfer score matrix, thereby obtaining the probability that the two labels are most likely to be adjacent.

D. Knowledge Fusion

After knowledge extraction, various entities are initially obtained, but there are a large number of equivalent instances, classes, and attributes in the entities, which is not conducive to the effective development of discipline inspection and supervision. Knowledge fusion is an effective way to solve the above problems. Knowledge fusion establishes the connection between heterogeneous ontologies and heterogeneous instances through the integration of heterogeneous data of knowledge from

different data sources so that heterogeneous knowledge graphs can be connected for mutual operation, which can not only improve the quality of the knowledge graph but also broaden the stock of knowledge [13]. The problem of ontology matching and instance matching in knowledge fusion can be attributed to several categories including terminology comparison based on natural language processing, matching based on ontology structure, and machine learning based on instance [14]. Knowledge fusion can enhance the robustness of the knowledge graph. It works better when there are various types of data sources.

E. Knowledge Processing

Through knowledge extraction and knowledge fusion, the ambiguity between entities is eliminated to the greatest extent, and basic factual expressions are obtained. However, the fact itself is not completely equal to knowledge. To form a structured knowledge network, processing processes including ontology construction, knowledge update, knowledge reasoning, and quality evaluation are also required [15]. Ontology construction is to summarize and extract knowledge in related fields to determine the terms that have been agreed upon and then describe the content through definitions. The relationship between the entity and its attributes is the concrete manifestation of the ontology, and the ontology has the function of generalization. Ontology construction consists of three stages, which are the calculation of parallel entity-relationship similarity, entity upper-lower relationship extraction, and ontology generation. For example, when the knowledge map receives “disciplinary payment collection”, “expulsion from the party”, and “private use of public vehicles”, it may be considered that the three entities are not much different. However, when calculating the similarity in the current period, it can be found that the first two become similar. After the ontology constructs the upper and lower relationship extraction, it is obtained that the first two belong to the punitive measure ontology, and the private use of the public car belongs to the problem situation ontology. Knowledge update is a key step to ensure the validity of the knowledge graph. The update of the concept layer and the update of the data layer can update the content of the knowledge graph including entities, relationships, and attribute values. When constructing the graph, after the unlabeled corpus predicts the entity through the trained model, the entity and relationship will be input into the graph database to complete the knowledge update. Knowledge reasoning can help discover new knowledge from existing knowledge, including inferring the current date by the punishment time. Quality assessment can quantify the credibility of knowledge, and judge whether the knowledge is credible through numerical values, to ensure the quality of the database and improve the accuracy of the content of the knowledge map [16].

F. Knowledge Storage

Through the above process, the relationship between the entity and the entity is obtained, and then an appropriate database needs to be selected for knowledge storage. Common storage methods include the storage method based on the relational database and the storage method of the graph database. Graph database has rich relationship display methods, so it is better than relational database when expressing many-to-many relational data. Currently, the graph databases with high activity are gStore [17], OrientDB [18], Neo4j, etc. Neo4j has the characteristics of efficient storage, rapid organization, and precise expression [19], and the query efficiency of the system itself is high. To

better achieve knowledge discovery and knowledge reasoning, in

this study, the graph database Neo4j was used to implement the storage of knowledge graphs. In the form of storage, nodes are used to represent entities in the knowledge graph, and edges are used to represent the relationship between entities. The relationship is directed, and the two sides

correspond to the start node and the end node. In the end, an easy-to-understand networked knowledge graph is formed to present knowledge. Neo4j adopts CQL (Cypher Query Language) query syntax. CQL has the characteristics of simple syntax and easy-to-understand structure, which also provides convenience for the subsequent use of py2neo for entity relationship import.

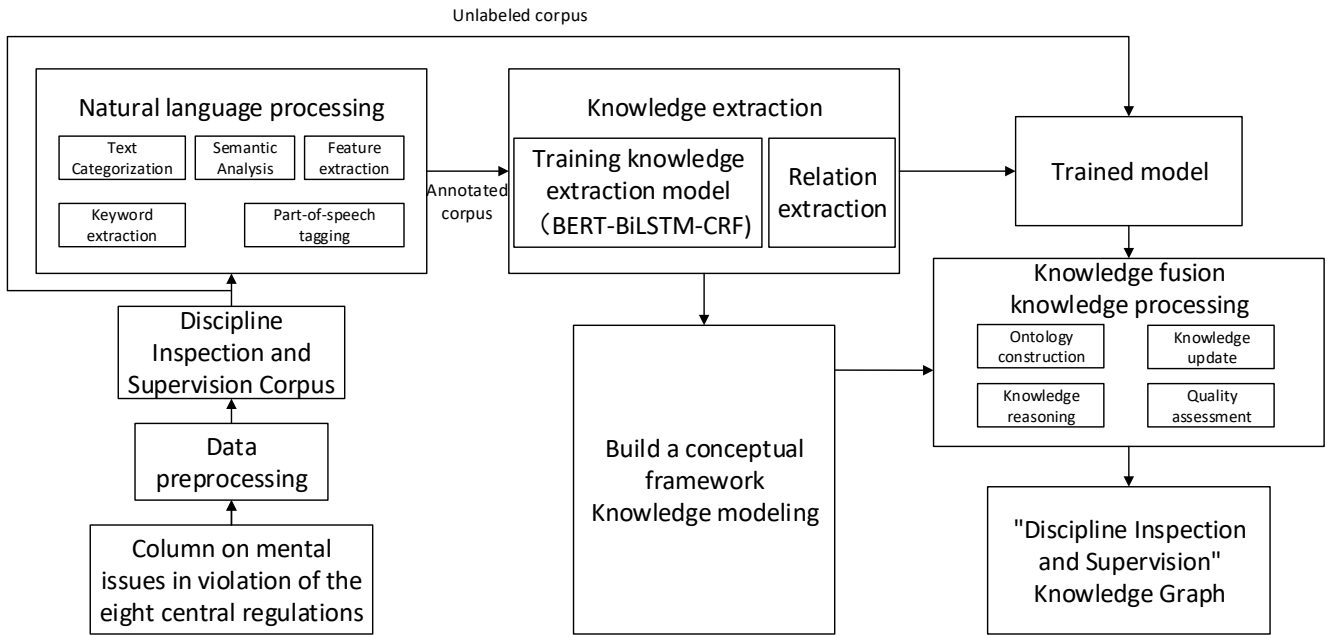


Fig. 3. Knowledge graph model of “discipline inspection and supervision”

III. IMPLEMENTATION OF "DISCIPLINE INSPECTION AND SUPERVISION" KNOWLEDGE GRAPH

For this map, the data sources, traverses, and reads the content of the notification were selected, and the notification corpus through processes including noise removal was obtained. After the corpus went through manual processing including word segmentation and sequence labeling, the knowledge extraction model training was carried out. The conceptual framework was constructed by the attribute relationship between the entities formed by the annotation and the extracted content, and the knowledge modeling was completed. Further, the unlabeled sentences were labeled through the trained model to complete the knowledge update. Finally, the tool py2neo was used to store the annotated corpus of the model in the Neo4J database to complete the knowledge storage. As shown in Fig. 3.

A. Dataset Acquisition

For this experiment, a column on mental issues that violate the eight central regulations of the Central Commission for Discipline Inspection and the State Supervision Commission’s website was obtained. As of January 9, 2021, a total of 10,031 pieces of data have been obtained, and the obtained notification file is in the .xls format. First, the file was converted into a data file, and the file was initially cleaned, with steps including removing duplicate files, removing summary sentences of the paper, and removing information including the names of the editor and the reviewer of the paper.

The notification had a fixed format, and the style of each paragraph was relatively uniform. Therefore, the clauses were marked with a period, and long parallel sentences and example sentences were divided into short sentences. Some

short sentences were shown in Fig. 4. Finally, 9,396 valid sentences were obtained, of which 5,259 were used as the training set, 2,178 were used as the verification set, and 1,959 were used as the test set.

1	杭州市司法局人民参与和促进法治处原处长施建良违规收受礼品礼金、违规接受管理服务对象宴请问题。
2	2015年至2019年，施建良在担任公证管理处（司法鉴定管理处）处长期间，多次收受6家司法鉴定机构相关负责人所送的礼品、礼金、消费卡，共计价值人民币（下同）51000元；
3	多次接受管理服务对象的宴请，收受土特产；将本人多次在滨江区某饭店消费的应当由个人支付的费用2770元，由下属单位支付、报销。
4	施建良还存在其他违纪违法问题。
5	2020年5月，施建良受到留党察看两年和政务撤职处分，降为四级调研员，违纪所得予以收缴。

Fig. 4. Part of the corpus

B. Chapter References

The sequence annotation of the dataset can indicate the type of word segmentation and the length of the word segmentation and other information. In this study, the BIO labeling system was used to label regions, organizations, departments, tasks, positions, problems, punishment time, and punishment measures. “B-X” indicates that the segment where this element was located was of type X and was at the beginning of this element, “I-X” indicates that the segment where this element was located was of type X and this element was in the middle of the segment, and “O” indicates other types. In this study, the Chinese word segmentation label was customized for classification, as shown in Table 1, and the text label examples are shown in Table 2.

TABLE I. DEFINITION OF WORD SEGMENTATION

Label	Note
LOC	Area
ORG	Organization
DEP	Department
PER	Person
JOB	Job title
PRB	Problem status
DAT	Punishment time
PUN	Punishment

TABLE II. DEFINITION OF WORD SEGMENTATION

Text	Sequence labeling	Text	Sequence labeling	Text	Sequence labeling
Xing	B-LOC	An	I-LOC	Meng	I-LOC
Ke	B-LOC	You	I-LOC	Qian	I-LOC
Qi	I-LOC	Wei	B-ORG	Sheng	I-ORG
Yuan	I-ORG	Yuan	B-JOB	Chang	I-JOB
Jin	B-PER	Jie	I-PER	Hua	I-PER
Wei	O	Gui	O	Shi	B-PRB
Yong	I-PRB	Gong	I-PRB	Che	I-PRB
Deng	O	Wen	O	Ti	O

C. Experiment Environment

Keras was used for model building and Python was used for programming. The CPU was Intel Xeon Gold6139. The GPU was GTX 2080Ti, and the video memory was 11G.

D. Model Implementation

To correctly obtain entities from unlabeled data, comprehensively utilize discipline inspection and supervision corpus, in this study, a BiLSTM+CRF model with BERT was implemented. The model can not only use BERT to capture the two-way relationship of word vectors in sentences but also use the bidirectional long and short-term memory network (BiLSTM) to extract the characteristics of text information. Finally, the constraint rules were automatically learned through CRF. The model structure is shown in Fig. 5.

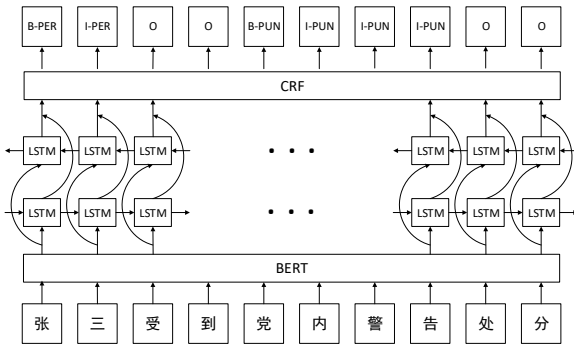


Fig. 5. Schematic diagram of model structure

The first layer is the input layer. The input layer takes characters as the unit. After two self-supervised tasks, masked language model (MLM) and next sentence prediction (NSP), the obtained input vector is passed to the BiLSTM layer to obtain the inherent characteristics of the characters. In this layer, the sequential sequence of the input

vector and the reverse sequence of the input vector are the inputs of the forward LSTM and the backward LSTM, respectively, so that a hidden state sequence is obtained. After inputting the hidden state sequence into the CRF and labeling the sentence, the sequence labeling result with the best accuracy is obtained by learning the constraint rules.

E. Model Performance Analysis

In this study, Keras was used for model building. After repeated comparison experiments, the optimal configuration of the model was finally determined. The maximum length of the input text MAX_SEQ_LEN was 128, the BATCH_SIZE of model training was 32, and there was a total of 10 rounds of training. To prevent overfitting, the dropout was set to 0.2, recurrent_dropout was set to 0.2, the optimizer was adam, the training learning rate was 0.001, and the loss function was crf_loss. Using the early stopping mechanism, early stopping occurred when the model was trained to the 7th round, and the optimal solution was obtained. Accuracy rate, recall rate, F1 value (harmonized average of precision rate and recall rate) were taken as the evaluation criteria of experimental results. In this study, three models BiLSTM, BiLSTM+CRF, and BERT+BiLSTM+CRF were used to annotate the discipline inspection and supervision corpus, which were trained for 296 rounds, 171 rounds, and 7 rounds respectively.

TABLE III. MODEL PERFORMANCE ANALYSIS

Models	Accuracy	Recall rate	F1 value
BiLSTM	0.488,6	0.550,0	0.517,5
BiLSTM+CRF	0.536,9	0.559,0	0.531,8
BERT+BiLSTM+CRF	0.685,2	0.717,2	0.700,9

It can be seen that the effect of using the combined model is better than the effect of using a single model. The addition of CRF effectively compensates for the irrationality and paranoia caused by BiLSTM in labeling. The effect of using BERT is better than the effect of not using BERT. The direct use of BERT's feature representation as a word embedding feature and its large number of pre-training models result in a higher F1 value. The experiment proved that the model proposed in this paper can improve the recognition effect of labeling discipline inspection and supervision corpus. Compared with the BiLSTM+CRF model, its accuracy, recall, and F1 value are higher by 14.83%, 15.82%, and 16.91%, respectively.

F. Knowledge Graph Generation

In this study, the py2neo tool was used to store the sentences predicted by the model in the "discipline inspection and supervision" knowledge graph database. First, the by Graph class passes the hostname, port number, username, and password to Neo4j for verification. After verification is successful, pre-set labels are defined as region, organization, department, task, position, problem, situation, punishment time, and punishment measures, respectively. The words obtained in the sentence are used to establish a neo4j node. Finally, a triple group is established through the relationship between words in each sentence, including "organization - belonging to - region", "person - appearance - problem", "person - being punished - punishment measure", etc. The part of the constructed knowledge graph is shown in Fig. 6, where the knowledge graph contains 27,689 entities and 14,909 associations, as shown in Table 4.

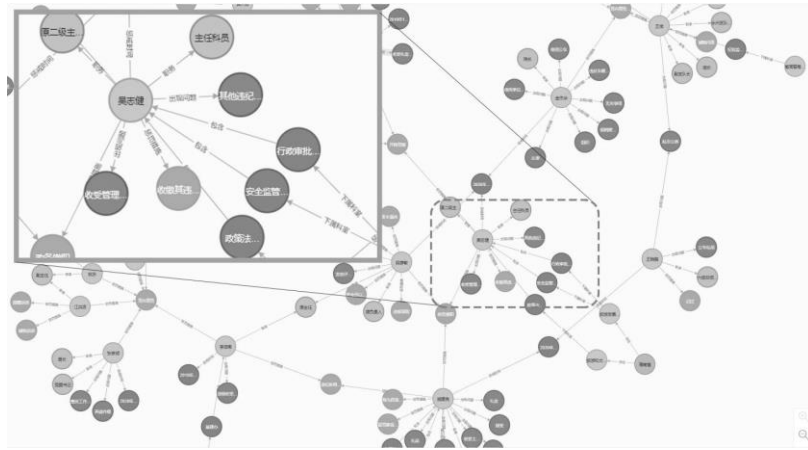


Fig. 6. Partial display of the knowledge map of discipline inspection and supervision

TABLE IV. ENTITY STATISTICS

Entity	Number of entities
Area	3684
Organization	3108
Department	751
Person	7652
Job title	4288
Problem status	3105
Penalty time	961
Punishment	4140

IV. CONCLUSION

This paper takes the construction of the “discipline inspection and supervision” knowledge map as an example, and introduces the whole process of the construction of the knowledge map based on the “column on mental issues in violation of the eight central regulations”. Through the practical processes including acquiring datasets, knowledge modeling, knowledge extraction, knowledge processing, and knowledge storage, the work from corpus to neo4j database storage was completed, and finally, an easy-to-use knowledge graph was generated. It provides a data source for knowledge base question answering and knowledge reasoning.

There are still some shortcomings in this study. First, due to the heavy workload of the manual sequence labeling project, there is a subjective style of labeling for everyone, which leads to the low quality of the dataset labeling, and the accuracy can be greatly improved when the labeling quality is improved; second, the experimental dataset in this study is the column content of the discipline inspection and supervision report. If the data sources can be appropriately expanded, it will be more conducive to discovering knowledge connections, thereby bringing more effective and comprehensive support to the discipline inspection and supervision work.

REFERENCES

[1] P. Zhou, Y. Hu, and S. Miao, "Research on Construction and Application of Knowledge Graph in Public Security Field," presented at the The 8th China Conference on Command and Control, BeiJing China, 2020.

[2] Y. Lei, J. Li, Y. Zeng, J. Wu, and X. Wang, "Construction of Military Knowledge Map Based on Multi-source Data Fusion," presented at the The 8th China Conference on Command and Control, BeiJing China, 2020.

[3] A. Zou, "Construction of Knowledge Graph Based on law," University Of Electronic Science And Technology Of China, Chengdu, 2019.

[4] K. Xue, "Research on the Construction and Application of Knowledge Graph in the Military Field," Dalian University of Technology, Dalian, 2020.

[5] X. Gao, "The construction of the legal knowledge graph of the application of artificial intelligence in civil justice—Based on the essential factual civil adjudication theory," *Law and Social Development*, vol. 24, no. 06, pp. 66-80, 2018.

[6] S. Liu, P. Chen, and J. Meng, "Practice and exploration on the knowledge graph of forging a strong sense of Chinese national community in cyberspace," *Computer Era*, no. 02, pp. 105-108, 2021.

[7] M. Liu, Y. Wang, Y. Wu, Y. Zang, and J. Liang, "Ontology Based Knowledge Modeling and Analysis of Military Equipment," *Computer and Modernization*, no. 01, pp. 76-80, 2021.

[8] W. Liu, "Research on the construction of knowledge graph of explosion-proof security inspection," *People's Public Security University of China*, 2020.

[9] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[10] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez et al., "Attention is all you need," *arXiv preprint arXiv:1706.03762*, 2017.

[11] J. Li, P. Wang, and Q. Han, "Computer Engineering & Science," *Computer Engineering & Science*, vol. 41, no. 04, pp. 713-718, 2019.

[12] L. Lli and Y. Guo, "Biomedical Named Entity Recognition with CNN-BLSTM-CRF," *Journal of Chinese Information Processing*, vol. 32, no. 01, pp. 166-122, 2018.

[13] Q. Liu, Y. Li, H. Duan, Y. Liu, and Z. Qin, "Knowledge Graph Construction Techniques: Taxonomy, Survey and Future Directions," *Journal of Computer Research and Development*, vol. 53, no. 03, pp. 582-600, 2016.

[14] T. Hang, J. Feng, and J. Lu, "Knowledge Graph Construction Techniques: Taxonomy, Survey and Future Directions," *Computer Science*, vol. 48, no. 02, pp. 175-189, 2021.

[15] Y. Shen and Y. Wang, "Knowledge Mapping of Public Crisis Events Based on Internet News Corpus," *Information Science*, vol. 39, no. 01, pp. 72-80, 2021.

[16] R. Yin, and Z. Yao, "Research on Data Quality Assessment Framework for Multidisciplinary Standards," *Standard Science*, no. 01, pp. 92-95, 2020.

[17] X. Zheng, "Analysis of the Impact of Consumer Behavior on User Consumption based on Gstore Data," *Journal of Anyang Institute of Technology*, vol. 19, no. 06, pp. 46-50+109, 2020.

[18] Y. Feng, "Design and Implementation of the Relationship Among the Characters Based on Knowledge Graph," 2016.

[19] M. Lu, M. Ye, and G. Zhang, "Construction and Implementation of Knowledge Map of Audit Data—Based on the Neo4j Graph Database," *China Computer & Communication*, vol. 33, no. 01, pp. 154-157, 2021.