

# Construction and query of power equipment knowledge map based on graph database

Xin Xuan<sup>a,\*</sup>, Qian Zheng<sup>a</sup>, Renzhe Xia<sup>a</sup>, Ruijie Wang<sup>a</sup>

<sup>a</sup>Beijing CHINA-POWER Information Technology Co., Ltd. State Grid Information & Telecommunication Group, Beijing 100085, China.

## ABSTRACT

Power equipment knowledge base is an important part of the digital transformation of power grid equipment management, but the structure of power equipment knowledge is complex, and its operation and maintenance involves equipment technical standards, general systems, product information, equipment encyclopedia and so on. In order to improve the utilization rate of data and mine the related information between devices, it needs to be processed and stored uniformly. In view of the above content, this paper introduces a construction and query method of power equipment knowledge map based on graph database. With the help of knowledge map technology, the text data of power equipment is processed, and the query method of power equipment knowledge is specified to complete the construction and query of power device knowledge map based on graph database. It realizes the standardized storage of power equipment knowledge, greatly improves the operation performance of complex associated data, and realizes the rapid query of power equipment technical standards, general systems, product information, equipment encyclopedia and other information. Experiments show that the model has a better recognition effect for entities with more times and stronger regularity, and achieves a more efficient and convenient effect in practical application.

**Keywords:** Power equipment knowledge, Graph database, Knowledge map, Graph storage, Graph query

## 1. INTRODUCTION

When operating on the first-line site, it is relatively inconvenient to obtain equipment data on the operation site. At present, most of the equipment data and standard systems are still stored in the form of traditional paper or electronic documents, which can not meet the real-time query application needs in the field operation process. In each link, the power grid equipment knowledge has not been connected, and the equipment knowledge can not be effectively utilized.

Digital construction is the key work of the State Grid Corporation of China, but at present, the field operation of the equipment inspection team lacks effective digital information support. With the continuous development of artificial intelligence technology, the process of intelligent processing of power equipment knowledge is further accelerated.

Artificial intelligence can strongly support the transformation and upgrading of traditional infrastructure of power grid enterprises, and provide support for data innovation drive and the emergence of new technologies, new models and new formats. "Digital new infrastructure" provides another strategic choice and development opportunity for power grid enterprises, which is an important starting point for the transformation and upgrading of infrastructure and core business of power grid enterprises, as well as an important opportunity for the comprehensive deepening of business transformation, mainly reflected in: accelerating the process of digital transformation.

\*e-mail: snowman1982@hotmail.com

Knowledge plays an increasingly important role in promoting social progress and development. Knowledge will be a major resource for development. "Knowledge Graph" was first proposed by Google in 2012 to enhance the performance of Google's search engine. The essence of knowledge mapping is a storage method, which stores the relationship between entities, and this storage method gives the application the ability of semantic recognition on this basis. With the deepening of research, knowledge mapping has been applied to medical treatment<sup>1-2</sup>, chemical industry<sup>3</sup>, electric power<sup>4-9</sup> and other fields. Knowledge mapping has played an important role in the following applications: intelligent search<sup>10-14</sup>, intelligent question and answer<sup>15-17</sup>, recommendation system<sup>18</sup> and so on.

At present, the main storage methods of power equipment knowledge are relational database and the storage based on extensible markup language (XML), which have the following problems: the efficiency of information retrieval is reduced when the amount of power equipment knowledge is large, there are great differences in the storage structure of power equipment information in different provinces and cities, and it is difficult to integrate them; Due to the low degree of structure, it is difficult to directly carry out in-depth data mining and fault analysis. The extraction of free text content in power equipment knowledge is highly professional and difficult to popularize.

This paper introduces a method of power equipment knowledge mapping construction and query based on graph database, and applies knowledge mapping technology to the research of power equipment knowledge storage and query. In this paper, the model based on BERT-BiLSTM-CRF<sup>19</sup> is used to extract power professional knowledge from power equipment information, as shown in Figure 1. The knowledge map of power equipment is constructed, and the storage and query based on Neo4j map database are completed. Compared with the traditional storage method, the rapid structuring of power equipment knowledge is realized, which provides strong support for the intelligent operation of front-line power equipment.

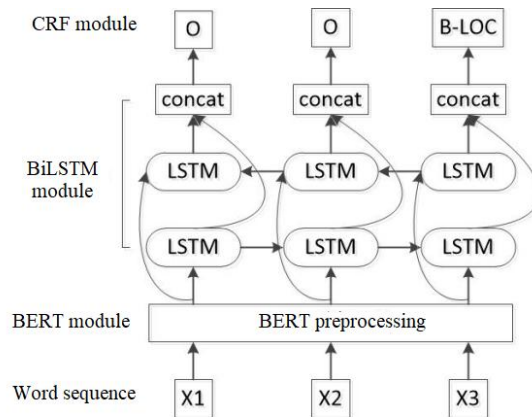


Fig. 1. BERT-BiLSTM-CRF Model Architecture

## 2. RELATED CONCEPTS

### 2.1 Knowledge map

Knowledge mapping is composed of entities, the relationship between entities and entities, and the attributes of entities. The knowledge graph is composed of pieces of knowledge, that is, SPO triples (subject-predicate-object).

Knowledge mapping is to represent knowledge, convert knowledge that people can understand into the form of mapping, so that machines can also understand natural language. The concept of knowledge map can be traced back to the semantic network proposed in the 1950s and 1960s. Semantic network is a form of knowledge representation, which is composed of interconnected nodes and edges. Nodes represent concepts or objects, and edges represent relationships between them.

The construction of knowledge mapping starts from the most original data (including structured data, semi-structured data and unstructured data), and adopts the technical means in the fields of natural language processing and data mining

to extract knowledge facts from the original database and store them into the knowledge base. This process includes three main processes: knowledge extraction, knowledge fusion and knowledge processing. Each update iteration contains these three phases.

## 2.2 Graph database

Through the steps of knowledge extraction and knowledge fusion in the process of building knowledge map, the required knowledge is obtained, and then the knowledge needs to be persisted for use. At present, in the knowledge map, the storage form of knowledge map is divided into two types according to the storage type: storage based on table structure (RDF) and storage based on graph structure (graph database). RDF (resource description framework), or Resource Description Framework, is a standard data model developed by W3C (World Wide Web Consortium) for describing entities/resources. In addition, relational databases can also store knowledge maps, which are actually composed of triples. Relational databases pay more attention to the internal attributes of entities, and the relationships between entities are usually realized by foreign keys. The relational database based on the table structure often requires time-consuming join operations. For most of the existing terabyte-level data, the relational database is often unable to meet the speed requirements, while the graph database technology makes up for this shortcoming, showing amazing performance when dealing with massive data, and plays an important role in the knowledge graph storage.

Graph database is widely used in various fields. It is a kind of non-relational database, which stores a graph structure. Graph is a complex nonlinear structure. Graph database consists of two important parts: node and relation, and each entity is a node. Nodes can have many attributes, and nodes are connected by edges (relationships), which are directional. In the graph structure, the relationship between nodes does not have only one direct predecessor and direct successor as in the linear structure, nor does it contain only hierarchical relationships as in the tree structure. The relationship between nodes is arbitrary, and any two data elements in the graph may be related.

Unlike relational databases, graph database relationships are not represented using foreign keys, but rather using edges for relational representation. For graph database, it can support knowledge storage and query with large amount of data and complex association. At present, the mainstream graph databases are Neo4j, Orient DB, Microsoft Azure, Cosmos DB and so on. There are many differences between graph database and RDF, as shown in Table 1:

Unlike relational databases, graph database relationships are not represented using foreign keys, but rather using edges for relational representation. For graph database, it can support knowledge storage and query with large amount of data and complex association. At present, the mainstream graph databases are Neo4j, Orient DB, Microsoft Azure, Cosmos DB and so on. There are many differences between graph database and RDF, as shown in Table 1:

Table 1. Comparison between RDF and graph database.

RDF	Graph database
Derived from metadata	From graph theory
Nodes and relationships cannot have attributes	Both nodes and relationships can have attributes.
Standard Inference Engine	No standard inference engine
Easy to publish and share data	Efficient query and search

## 2.3 Word2Vec

In terms of language processing, the most granular unit is the word. Sentences are formed from words, and then paragraphs and chapters are formed from sentences. Therefore, to deal with natural language, we must first deal with words. There is a task to determine whether a word is a verb or a noun. If we use the method of machine learning, we already have a series of labeled samples  $(x, y)$ , where  $x$  represents the word and  $y$  represents the part of speech corresponding to the word. We want to identify the part of speech of the unlabeled word. The idea is to build a mapping of  $f(x) \rightarrow y$ , that is, to create a neural network whose input is the word and its output is the corresponding part of speech of the word. Train this model to get the ability to identify the part of speech of words. But there is a problem here, usually the input words are natural language, which is an abstract summary of human beings and can not be understood

by machines, so we need to find ways to convert natural language into a form that machines can understand, that is, numerical form. This way of converting natural language into numerical form is called word embedding. Word 2vec is a kind of word embedding.

The following is a brief introduction to the implementation process of Word2vec. Taking Chinese as an example, a high-frequency dictionary is constructed according to all the text data in the data set, that is, the sentence is segmented at the character level to form a dictionary. For example, given the sentence:

“This, this, this, is, is, is, power, power, power, voltage, voltage, voltage, transformer, transformer, transformer.” The dictionary constructed is shown in the Table 2:

Table 2. High frequency dictionary.

Serial number	Word
1	This
2	is
3	power
4	voltage
5	transformer

Each word corresponds to an index, indicating the arrangement position of the current word. Take "Power voltage transformer" as an example to perform the Word2vec operation steps:

that length of the dictionary in Table 2 is 5, and an all-zero vector  $VEC = [0, 0, 0, 0, 0]$  with a dimension of 5 is initialize;

1. Obtain the index of the word to be represented in the dictionary. "power" is 3, "voltage" is 4, and "transformer" is 5.
2. The value of the corresponding index position in the vector VEC is set to 1, and the other positions are set to 0;
3. Then the "Power voltage transformer" can be represented by a vector as:  $[0, 0, 1, 0, 0]$ ,  $[0, 0, 0, 1, 0]$ ,  $[0, 0, 0, 0, 1]$ ;
4. Through the above conversion, the "Power voltage transformer" is converted into a numerical form that the machine can recognize, and the neural network model can be trained by taking the data in the numerical form as the input and the category as the output. The final model can get the type of the word by inputting the word.

## 2.4 Search engine

### (1) Full-text search engine

Google and Yahoo are the representatives in foreign countries, and Baidu and Peking University Skynet are the representatives in China. The principle of this kind of search engine is to store all the content on the Internet into its own data, then retrieve the records matching the user's query conditions, and then display them to the user in a predetermined order.

### (2) Directory search engine

This kind of search engine classifies the list of website links according to the directory, and users can find the information they need only by the classified directory, the most representative of which is Yahoo. Sohu and Sina in China.

## 2.5 Semantic search

Tim Berners-Lee, the father of the World Wide Web, explained that "the essence of semantic search is to use mathematics to get rid of the guesses and approximations used in today's search, and to introduce a clear understanding of the meaning of words and how they relate to what we find in search engine input boxes".

With the emergence of the concept of Semantic Web, more and more open linked data and user-generated content are published on the Internet, and the Internet has been transformed into a data network that contains a large number of entities and relationships between entities. In this context, Google proposed the concept of knowledge graph in May 2012, which aims to describe the relationship between various entities in the real world, so as to improve search results.

Following that, Sogou put forward "knowledge cube", Microsoft put forward "Probase" and Baidu put forward "intimate".

The working principle of semantic search engine is that it not only pays attention to the user input content, but also pays attention to the meaning expressed by the user input content, accurately understands the real intention of the user, searches with this semantic information, and can return more accurate search results to the user. Compared with the traditional keyword-based search engine, there is a great progress.

### **3. CONSTRUCTION OF CORPUS FOR POWER EQUIPMENT KNOWLEDGE ANNOTATION**

#### **3.1 Ontology library**

Ontology is the base of knowledge base. Ontology is at the conceptual level, which is similar to the class in programming language. Instance is a concrete presentation of ontology, which is similar to the instantiation in programming language.

When establishing a knowledge base, an ontology base should be established first. For example: I establish an ontology base, in which only one class is established: equipment, which has attributes such as service life, date of manufacture, model, etc., and other relationships are not considered for the time being; supplementary example data: transformer, transmission line, transformer, etc.

Therefore, a knowledge base is established to store the power equipment ontology and the instance data of the power equipment.

#### **3.2 Construction of Corpus for Power Equipment Knowledge Annotation**

In this paper, a corpus of power equipment knowledge annotation is built, which contains 4000 power equipment knowledge related data from provincial power grid companies, as well as unstructured or semi-structured data related to power equipment on the Internet. Due to the differences in the structure of the data obtained from various sources, after data cleaning work such as desensitization and non-text content processing, the regular matching method based on rules is used to extract the power equipment knowledge text of the data and store it in MYSQL database. We have also worked with power industry experts to develop a set of marking rules, using the YEDDA marking tool for manual marking. The marking personnel are all composed of power industry workers. The marking work is carried out in a back-to-back manner, and the final review is completed by power industry experts.

### **4. POWER EQUIPMENT KNOWLEDGE MAP CONSTRUCTION**

The knowledge map takes the power equipment knowledge as the center, the equipment, experiment, experimental environment and device as the main nodes, the voltage, current level, power size, functional characteristics, location, experimental method and experimental duration as the main attributes, and the experimental equipment relationship, experimental device relationship and experimental environment relationship as the main relationships. The knowledge map construction process mainly includes the following three steps:

#### **1) Knowledge abstraction**

We abstract entities, attributes and the relationship between entities from unstructured, semi-structured and structured power domain data, and form ontological power domain knowledge. For unstructured text data, we abstract the entities, relationships, and attributes in the data. The following technical practices need to be done: one is domain entity abstraction, also known as entity recognition, where the entities are usually people, organizations, places, etc.; the other is relationship abstraction, that is, the relationship between entities, which uses certain technical means to abstract the relationship information. The third is the attribute abstraction, that is, the abstraction of the attribute information of the entity, which is similar to the relationship. The relationship reflects the external relationship of the entity, and the attribute reflects the internal characteristics of the entity.

#### **2) Knowledge fusion**

Knowledge fusion is the process of integrating the knowledge in multiple knowledge bases to form a knowledge base. After acquiring new knowledge, it is necessary to integrate and eliminate contradictions and ambiguities. For example, some entities will have multiple forms of expression, and a certain form of expression is for multiple different entities. Different knowledge bases, for the same entity, some knowledge bases may focus on the description of a certain aspect of itself, and some knowledge bases may focus on the description of the relationship between the entity and other entities. The purpose of knowledge fusion is to integrate them so as to obtain a complete description of the entity.

### 3) Knowledge processing

For the new knowledge after fusion and reasoning, the qualified part can be added to the knowledge base after quality evaluation (some of them need to be screened manually), so as to ensure the quality of the knowledge base, and the purpose is to obtain the knowledge map data that meets the requirements.

The detailed construction process of knowledge map is shown in Figure 2.

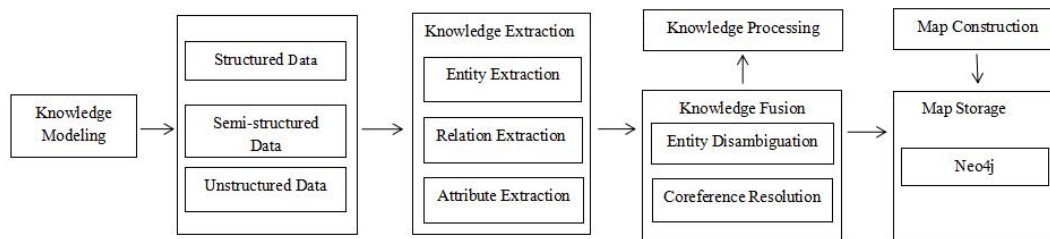


Fig. 2. Knowledge mapping construction process

## 5. QUERY BASED ON GRAPH DATABASE

### 5.1 Necessity of introducing knowledge map

The traditional keyword-based search system has the problem of low precision and recall, which is due to the fact that the machine can not understand the meaning of the user's input content, resulting in the search answer returned by the machine often can not meet the user's needs. Based on the knowledge map of power equipment constructed above, the power equipment knowledge is stored in the graph database Neo4j, and the traditional keyword-based power equipment knowledge search system is improved through word segmentation of user input sentences, entity relationship extraction, and graph database query technology.

As the core of knowledge map, the establishment process of knowledge base is a continuous process. The content of knowledge base is not immutable, but needs to be updated and integrated iteratively. Whenever there is a user searching, through the intelligent semantic understanding of the query sentence input by the user, the stored content is automatically retrieved and matched in the knowledge base, and the results are presented in a visual way.

The disadvantages of the traditional search model are that it is difficult to understand the user's intention, to match accurately, and to provide personalized services. The introduction of knowledge graph can solve the following reasons: knowledge graph can express the association between query and answer, and provide interpretable basis for search.

### 5.2 Semantic search method based on knowledge map

This paper combines the traditional keyword search method and the entity query method based on knowledge map, first uses the traditional keywords to roughly determine the search scope, and then uses the knowledge map subgraph query to accurately search the semantic. The basic process comprises the following steps of: 1) identifying keywords of the content input by a user, and positioning a candidate knowledge map subgraph matched with the search content, thereby accelerating the efficiency of entity search in the knowledge map; 2) identifying an entity of the content input by the user, generating a Neo4j Cypher query statement, and performing entity search on the positioned knowledge map subgraph; 3) discover the relationship between entities in the user's query content through the graph database query, and then

understand the user's search intention; 4) sort the search results (by the importance in the knowledge graph structure, by the popularity of entities, and by the relevance to the query). The semantic search process based on knowledge mapping is shown in Figure 3.

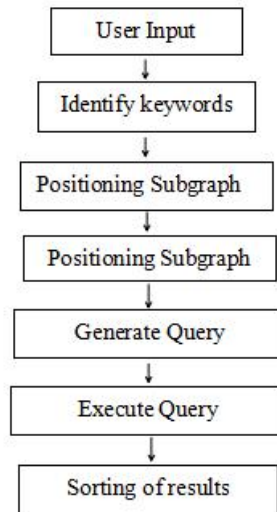


Fig. 3. Semantic Search Process Based on Knowledge Map

The traditional keyword search method, because the machine can not understand the user's search intention, so the search results usually have a large error, while the semantic search method based on knowledge map, when querying, because it needs to traverse the entire knowledge map, it will take a lot of time. Combining the advantages of the two, in the initial stage of search, the keyword search is used to quickly locate the target search area, and then the knowledge map subgraph query is carried out in this area, which not only ensures the accuracy of the search results, but also improves the overall search efficiency.

## 6. SUMMARY

Aiming at the low precision and recall rate of traditional keyword search technology in the application of power industry, this paper introduces a method of knowledge map construction and query of power equipment based on graph database, which constructs the knowledge map of power equipment, stores the knowledge of power equipment into the graph database, and combines the traditional keyword search and graph database based on knowledge map. The accuracy and efficiency of power equipment knowledge search are improved.

## REFERENCES

- [1] Xie, Y. L., Cai, P. Q., Jiang, W. and Li, K., Storage Method of Electronic Medical Record Based on Graph Database [J]. Information Technology and Informatization. (08):134-137 (2021).
- [2] Zhao, X. W., Research on Question Answering System Based on Knowledge Mapping in Medical Domain [D]. Harbin University of Science and Technology (2021).
- [3] Zeng, W. G., Research on Knowledge Mapping of Chemical Safety Based on Neo4j[J], Heilongjiang Science, 12(16):17-19 (2021).
- [4] Gong, Y., Li, B. W., Power equipment fault knowledge base construction method based on knowledge map[J]. Reliability and Environmental Test of Electronic Products, 39(04):72-77 (2021).
- [5] Ji, Y., Xie, D., Method for constructing semantic search system in electric power field[J]. Computer system application, 25(04):91-96 (2016).

- [6] Zhao, S., Qi, X. M., Research on Recommendation Search Technology of Electrical Equipment Based on Knowledge Map, *Electronic devices*, 44(01):182-187 (2021).
- [7] Fu, X., Guo Y., Design of Power Grid Operation Monitoring and Analysis System Based on Knowledge Mapping Technology[J]. *Power Supply and Power Consumption*, 38(07):45-50 (2021).
- [8] Gao, H. X., Miao, L., Overview of Knowledge Mapping and Its Application in Power System[J]. *Guangdong Electric Power*, 33(09):66-76 (2022).
- [9] Song, H. Y., Research and Application of Power System Knowledge Mapping Based on Graph Database[D]. University of Chinese Academy of Sciences (2021).
- [10] Liu, Y. F., Research and Application of Search Engine Technology Based on Knowledge Map[J]. *Wireless Internet technology*, 18(06):95-96 (2021).
- [11] Liu, J. Z., Wang, Y., MOOC Platform Resource Retrieval Engine Based on Knowledge Map[J]. *Modern Vocational Education*, (24):60-63 (2021).
- [12] Wang, M., Wang, J. T., Active Search of Knowledge Map Based on Human-Computer Hybrid[J]. *Computer Research and Development*, 57(12):2501-2513 (2020).
- [13] Ruan, G. C., Fan, Y. H., A Review on the Application of Mapping Knowledge Domains in Entity Retrieval[J]. *Library and information work*, 64(14):126-135 (2020).
- [14] Zhou, J., Sun, X. M., Application of knowledge mapping in semantic information search accuracy[J]. *Computer and Digital Engineering*, 48(06):1445-1449 (2020).
- [15] Zhang, Q., Research and Application on Key Technologies of Question Answering System Based on Knowledge Mapping [D]. University of Chinese Academy of Sciences (2021).
- [16] Xu, M. T., Multi-round Question Answering System Based on Knowledge Map[D]. Nanjing University of Posts and Telecommunications (2020).
- [17] Wang, Z. Y., Yu, Q., Wang, N., Survey of Intelligent Question Answering Based on Knowledge Mapping[D]. *Computer Engineering and Application*, 56(23):1-11 (2020).
- [18] Qin, C., Zhuang, F. Z., Zhang, Q., A Survey of Knowledge Mapping Based Recommender System[J]. *Science in China*, 50(07):937-956 (2020).
- [19] Xie, T., Yang, J. A., Liu, H., Chinese entity recognition based on BERT-BiLSTM-CRF model[J]. *Computer Systems & Applications*, 29(7): 48-55 (2020).