Research on Power Equipment System of Knowledge Graph under Electric Energy in Smart Grid

Jiangyi Hu¹, a, Nan Zhang², b, Yingnan Shang², b, Jun Yu², b, Wenqiang Zhang², b, *

¹Nanrui Group Co., Nanjing 210000, China
²Nari Technology Development Limited Company, Nanjing 210000, China

*Corresponding author: zhangwenqiang3@sgepri.sgcc.com.cn

Abstract. In order to realize technological innovation in the electric power field, the article uses artificial intelligence, big data analysis and mining, knowledge graphs, natural language processing and other technologies to construct Chinese professional dictionaries and knowledge graphs in the electric power field. And on this basis, the thesis puts forward the technical architecture of a comprehensive energy service support system based on the concept of "a picture of the power grid", a spatiotemporal data management platform and a graph database query application implementation mechanism. Finally, research and design the visualization of the hierarchical organization of the data, use the association relationship between the data to organize the data into multi-topic classification and hierarchical organization, and provide a friendly interactive man-machine interface.

1. Introduction

With the development of various businesses of the smart grid, China's smart grid has come to the forefront of the world in many aspects such as UHV, and the power automation system of the power grid plays an increasingly indispensable role in the safe operation of the smart grid. There are more and more various applications of power automation systems, and business knowledge is becoming more and more complex. Most relevant business personnel only understand partial business knowledge, but do not know the related business logic. Only a few expert-level personnel can understand the entire business process. Have a clear understanding. Therefore, when there is a problem with complex business logic, it is necessary to temporarily mobilize all business personnel to clarify the business logic relationship, and then it is possible to find the cause of the problem.

Current research on quality management of power equipment is usually based on relational databases. Literature designed a set of power equipment quality management system based on B/S architecture. The designed system adopts a unified standard for equipment management, which greatly reduces the workload of equipment management personnel. The literature combines the multi-layer CS software architecture with the orthogonal software architecture, and proposes a comprehensive management platform for the quality of power equipment based on the heterogeneous orthogonal architecture, which has strong compatibility and data Synchronization. In order to realize the real-time management of the operating status of power equipment, literature proposed a comprehensive management system for power equipment defects based on workflow, thereby improving the reliability of power equipment operation [1]. However, with the continuous deepening of big data applications in recent years, a series...
of technical bottlenecks have become increasingly prominent in the traditional relational database used in power grid automation systems to deal with the complex relationship problems of big data, such as: complex data relationship structure conversion, data retrieval and query Low efficiency, huge waste of storage space, difficulty in supporting large-scale concurrent access, and difficulty in database application expansion. To this end, this article uses knowledge extraction, knowledge expression and graph database technology, combined with graph computing-based knowledge graph and intelligent reasoning technology to form an intelligent graph computing platform, establish a power equipment knowledge base, and break through the barriers between equipment data and monitoring data, Carry out comprehensive management for the quality of power equipment, further improve the level of equipment quality management and operation and maintenance, and realize the comprehensive management of power equipment.

2. Applicability analysis of power hotspots of knowledge map

2.1. Overview of knowledge graph theory

The concept of knowledge graph was formally proposed by google in 2012. The main purpose is to improve the intelligence and efficiency of search engines. The knowledge graph is essentially a semantic network, with nodes representing entities or attributes, and edges representing various semantic relationships between entities and between entities and attributes. Among them, entities refer to objects or things that exist objectively in the real world and are distinguishable, such as China and peaches [2]. Attributes are information describing the characteristics of an entity, such as area and maturity. Relationship is the most important feature of the knowledge graph. Only by this can the interconnection of everything be realized, thereby supporting various applications such as semantic understanding and information retrieval. The knowledge graph construction technology mainly includes: knowledge extraction, knowledge fusion, knowledge representation, knowledge verification and knowledge reasoning. The construction process framework is shown in Figure 1.

![Knowledge graph construction process](image)

Figure 1. Knowledge graph construction process

2.2. Applicability analysis of power knowledge graph

As a major national energy support system, the power industry is widely distributed and has a complex structure. Massive data will be generated in all aspects of power generation and power services such as power generation, transmission, transformation, distribution, and power consumption. The power plant will generate data such as water level, temperature, and variable-speed wind; smart meters collect data every 15 minutes, and the interval for collecting data in an ideal state can be less than 1 s. In the transmission grid, basic information such as various switching signals, voltage, current, power, and transformer oil temperature can also be refreshed at least once in 1 s. Under such high frequency recording, a considerable amount of data will be formed in the background. The State Grid Corporation of China is currently operating 240 million smart meters, with an average annual data volume of about 200 TB; in addition, power companies will generate a large amount of talents, materials, power market information, capital operations, and collaborative office during operation and management. data.
3. Automatic construction of knowledge map of power equipment defects

3.1. The process of constructing the knowledge map of power equipment defects
Defect records of electrical equipment usually exist in the form of a single sentence, and the equipment parts, phenomena, and degrees of defects are generally recorded in natural language. Since the defect records of power equipment have their own characteristics, when constructing the knowledge map of power equipment defects, the following modifications are made on the basis of the general process of knowledge map construction. 1) In power equipment defects, as the attribute of the defective component, the defect phenomenon itself also has attributes such as the degree of defect, so in addition to extracting the relationship between entities and the relationship between entities and attributes, the relationship between attributes should also be extracted. 2) The power equipment defect knowledge graph belongs to a closed domain knowledge graph. The meaning of entity words is limited to the power field, and the power industry has clear terminology specifications. The entity ambiguity problem basically does not exist, and the entity disambiguation step is omitted. The above-mentioned modified knowledge graph construction process is shown in Figure 2. Except that the data integration method is basically the same as the general knowledge map, the rest of the steps need to be designed specifically, which will be highlighted below.

![Figure 2. The process of constructing a knowledge map of power equipment defects](image)

3.2. Entity/attribute extraction
The main task of entity/attribute extraction is to extract the words representing entities/attributes in the defect record corpus of electrical equipment, and perform part-of-speech tagging. Since entities and attributes can be exhaustively enumerated, they can be matched and extracted directly using a professional electric power dictionary. The specific steps are as follows. 1) Word segmentation. First, we segment the corpus of defects in electrical equipment. The word segmentation is based on a dictionary of commonly used words and a hidden Markov model, and a professional electric power dictionary is imported to assist in word segmentation to improve accuracy. 2) Word extraction. The words in the segmented corpus are retrieved one by one in the electric power professional dictionary. If the matching item can be retrieved, the entity/attribute represented by the word is extracted as the entity/attribute of the knowledge graph. 3) Filter word pairs. When vectorizing words, the words that
are adjacent in the sentence (neighboring word pairs), such as the "oil storage tank" and "rust" in the "main variable body oil storage tank rust", or words with similar contexts (appositions) Yes), such as the "oil conservator" in the "main transformer body oil conservator rusted" and the "oil conservator" in the "main transformer body oil conservator rust", both have a high cosine similarity. Obviously, the apposition pair is the synonym to be looked for, and the two words of the apposition pair can hardly appear in the same defect record. Therefore, word pairs that have appeared in the same defective record can be deleted, so that adjacent word pairs can be eliminated, and composition word pairs can be screened out [3].

3.3. Relation extraction
The main task of relationship extraction is to identify whether there are relationships between entities/attributes and the corresponding relationship types. The power equipment defect knowledge graph can be combined with the part of speech of the entity/attribute to limit the relationship, as shown in Table 1.

Table 1. "Type of relationship between entities/attributes

| Entity/attribute A part of speech | Entity/attribute B part of speech | Possible relationship between A and B |
|----------------------------------|----------------------------------|------------------------------------|
| En                               | En                               | A contains B/B contains A/no relation |
| En                               | Pv                               | A defect phenomenon is B/no relation |
| Pv                               | Pad                              | A defect degree is B/no relation |
| Pv                               | Pq                               | A defect quantification unit is B/no relation |

Before the relationship classification, the word pairs to be classified are formed first, that is, among all the pairwise combinations of words representing entities/attributes, the word pairs belonging to the 4 part-of-speech combinations in Table 1 are selected, and then each word pair is classified [4].

4. Knowledge graph modeling technology based on power equipment

4.1. Basic technology
The concept of establishing a schema based on the knowledge graph database of a graph database is a bit like a namespace or you can think of it as a directory in a file system. The difference is that there can be no more schema nesting under this schema. Objects such as tables, functions, etc. are stored under each node, edge, or attribute defined in the schema. There cannot be repeated object names under the same schema, but they can be repeated under different schemas. The open source graph database is characterized by supporting native graph storage and processing, and ACID transaction processing, but does not use schema. Therefore, in the enterprise data management scenario, it is difficult to control the data as a whole [5]. In addition, the open source graph database does not support the storage of real-time information, and the non-enterprise version is limited by data volume and query speed. Using schema is convenient for multiple business departments and multiple users to share a database, and at the same time, different data sources can be independent of each other, which facilitates the management of many objects and is more logical. For example, to design a complex system composed of many modules, different modules need to be independent, but storing the data of each module separately in an independent database cannot achieve the overall integration and unified modeling of the modules. Use the schema to classify the differences between the modules. Objects, and then control users with appropriate permissions, can logically and clearly construct a unified data model.

From the perspective of equipment management, set the equipment as a node, and at the same time set the equipment-related information and the topological connection relationship of the equipment in
the power grid, such as equipment type, equipment voltage level, power supply company using equipment, equipment manufacturer, equipment characteristics Parameters, equipment failures, defects, equipment installation sites and lines, etc. (see Figure 3) are set as nodes, and the relationship between the equipment node and other nodes, such as "voltage level", "installation at", "manufacturer" Set “Yes” and “Occurrence” as edges to construct a device-centric schema to facilitate the conversion of data from the relational data format of the table to the graph database format [6].

Figure 3. Construction design drawing of device-based graph database schema

4.2. Main functions
The application of the integrated power equipment quality management system based on the graph database is oriented to the entire industry chain of power equipment quality management, such as power equipment management, quality traceability, quality assessment, key equipment analysis, active maintenance, operation and maintenance, etc., to achieve power equipment as the core The closed-loop system application not only provides various knowledge searches for power equipment regulations, regulations, equipment manufacturers, service stations, power supply companies, etc., but also provides applications for power equipment knowledge graphs and power equipment graph databases for complex knowledge reasoning analysis. The established knowledge graph of power equipment covers the information of more than 2 million devices in 39 categories of the main network and distribution network of a certain provincial company. The total number of nodes in the graph database exceeds 2.2 million, and the total number of edges exceeds 25 million. The main interface of the power equipment quality management system based on the graph database is shown in Figure 4. The system overview records the running time of the equipment quality management system and the overview of the overall equipment health indicators in the database; the real-time health indicators dynamically update the current system health every 15s The percentage of equipment in the total number of equipment in the system; the situational awareness diagram shows the equipment in the power grid from six aspects: system reliability operation indicators, equipment health indicators, equipment quality indicators, equipment operating environment factors, equipment operating limits, and network topology robustness [7]. The dynamic results of situational awareness related indicators in operation; the pie chart records the percentages of healthy equipment, faulty equipment, and defective equipment in 39 types of equipment; the scroll bar dynamically displays the latest equipment defect or failure related information.
5. Conclusion
Based on the introduction of the characteristics and technical architecture of the graph computing platform, this paper presents the technical architecture and application modules of the integrated energy service technical support system for the active distribution network based on the graph computing platform, and proposes a concept based on the “one picture of the grid” Spatiotemporal data organization and management and massive data visualization technology, researched fast parallel computing technology based on graph computing and new graph computing mode; developed 10kV distribution transformer load/distributed photovoltaic power generation forecasting technology based on graph machine learning; discussed based on virtual power plant Conceptual distributed resource optimization control management mechanism; explored the power grid equipment quality cross-system full life cycle management technology based on knowledge graph. It is hoped that these new technologies can provide new technology options for the active power distribution network integrated energy service business that is currently booming.

References
[1] Wang, H. An error recognition method for power equipment defect records based on knowledge graph technology. Frontiers of Information Technology & Electronic Engineering, 20(11) (2019) 1564-1577.
[2] None. Assured quality management system. British Dental Journal, 223(7) (2017) 543.
[3] Patel, S. The global quality management system: improvement through systems thinking. Lancet, 374(9707) (2016) 2028.
[4] Zhang, S. Y., Gu, Y., Yi, G. D., & Wang, Z. L. A knowledge matching approach based on multi-classification radial basis function neural network for knowledge push system. Frontiers of Information Technology & Electronic Engineering, 21(7) (2020) 981-994.
[5] Ando Junior, O. H., Bretas, A. S., & Leborgne, R. C. Methodology for calculation and management for indicators of power quality energy. IEEE Latin America Transactions, 13(7)
[6] Leroy, E., & Marque, A. [audits of the quality management system and safety in radiotherapy: lessons learned and future prospects]. Cancer Radiotherapie, 20(6-7) (2016) 614-615.

[7] Liu, Y., Ni, L., & Fu, X. The roles of structure-based and function-based action knowledge in object recognition. Journal of Vision, 15(12) (2015) 235.