Research Article

Research on Regional Energy Visualization Based on Knowledge Graph

Zhongbo Li [ID] [1,2]

[1]*College of Energy Engineering, Zhejiang University, Hangzhou 310058, Zhejiang, China*
[2]*Beijing District Heating Group, Beijing 100028, China*

Correspondence should be addressed to Zhongbo Li; 11827107@zju.edu.cn

The equipment information in the regional energy system is difficult to retrieve and the amount of data is too large to realize intelligent application. In order to solve the problem, this paper proposes a regional energy visualization research method based on knowledge graph. Firstly, the collected regional energy information is classified. Secondly, the framework is defined for the unstructured data, which is difficult to be recognized by computer, and the hidden Markov model is used for word segmentation and labeling. Finally, the regional energy equipment information data are processed to construct the knowledge graph, the regional energy map structure information is represented by the ternary group graph structure, and the regional energy retrieval system is designed. The system can efficiently process and retrieve a large amount of natural language information in regional energy and realize the intelligence of regional energy equipment. It provides a strong guarantee for the efficient and stable operation of regional energy-related equipment and has a wide range of practical application and research value.

1. Introduction

District energy systems are local energy networks through which it is used to supply hot water, steam (district heating), cold water (district cooling), electricity (often called microgrids), or integrated supply to complex buildings. In the daily construction and operation of regional energy system, a large number of equipment parameters and operation data are accumulated, but the corresponding information is often idle in the system, which affects the data value discovery of regional energy system. With the increase of equipment information, how to retrieve equipment information efficiently is of great significance to the staff of regional energy system. However, the complexity of regional energy system equipment makes it difficult to accurately retrieve equipment information [1].

(1) The primary equipment information of the regional energy system is described in the form of natural language, which increases the difficulty for the computer to understand the equipment information, and the complexity of the equipment information causes great difficulties for the accurate retrieval of the information;

(2) The intelligence degree of the primary equipment information management is not enough and it cannot select useful information to utilization from a large number of monitoring data and historical data.

The continuous development of artificial intelligence technology, the Internet of Things, and other technologies provides a new direction for the intelligent management of regional energy system equipment information [2]. Natural language processing (NLP) is an intelligent and efficient text processing technology that can systematically analyze, understand, and extract key information from text data. As for the equipment information corpus of regional energy system, literature [3, 4] expresses the text by establishing semantic framework, but the semantic framework is difficult to adapt to the complex information situation of power equipment [5], and the framework relies on the definition of expert experience, so it is difficult to comprehensively

consider the complex expression mode of equipment of regional energy system. Literature [6, 7] uses machine learning algorithm to mine rules in corpus to represent features of corpus. However, the features selected by machine learning methods are basically limited to the occurrence of keywords [8] or the occurrence frequency of words [9]. These statistical features do not fully consider the internal logic of keywords in sentences, although they have certain regularity, they are easily limited to the textual features of defect records, which is not enough to explain. With the continuous development of intelligent regional energy system strategy, the equipment information of regional energy system increases exponentially, and the requirements for information storage and retrieval are higher and higher. As a kind of efficient database, knowledge graph can manage regional energy system equipment information effectively and provides a new way for regional energy equipment information management. Different from the traditional literature review [10–13], the literature knowledge graph can extract and screen structured knowledge sequences from a large number of literature data, then show the evolution process of research hotspots and map the cross-interaction between knowledge groups, so as to realize the mining of hot spots and emerging trends. At present, common mapping tools include CiteSpace [14], etc., which provides an efficient, repeatable, and rapidly applied analysis method for scholars and has application prospects in various fields.

Therefore, this paper tries to apply the knowledge graph method to the primary equipment information management of intelligent area energy system. Firstly, the data collected by the intelligent regional energy system information collection platform are classified. Secondly, considering that the unstructured data are difficult to be accurately recognized by the computer, the unstructured data extraction framework is defined, and the hidden Markov model is used to classify and annotate the collected data, providing data support for the construction of the primary equipment information knowledge map of the regional energy system. Finally, the knowledge graph of regional energy system is constructed on this basis, the regional energy system information is represented in the form of ternary group graph structure, and the primary equipment information retrieval system of regional energy system is designed. The system can effectively improve the retrieval efficiency of regional energy system equipment information and improve the intelligence level of regional energy system.

2. Regional Energy Data

2.1. Data Sources for Regional Energy Systems.
Data sources of regional energy system mainly include original data, internal data of each automation system in the regional energy system, fault error recording information, and data obtained from monitoring of the surrounding environment [15]. The primary equipment of the regional energy system is directly connected to the high voltage grid of the power system and participates in the transformation, transmission, and distribution of power energy. There are many kinds of primary equipment in regional energy system, including transformer, circuit breaker, disconnecting switch, and other related equipment. Transformer is the main equipment for the conversion of AC power energy in the power system. It can realize the conversion of power energy between different voltage levels, so as to facilitate the connection of the power system and optimize the transmission of power energy. Transformer data not only include voltage, current, active power, reactive power, phase, and other data, but also include transformer online monitoring data, chromatographic data information, and also include the proportion of dissolved gas in power transformer insulation oil and gas production data information [16]. These data parameters constitute the main components of power transformer data.

The circuit breaker is an important piece of equipment to connect or disconnect the equipment of the power system. It is equipped with relatively perfect arc extinguishing device and has the ability to disconnect the current of the power system. The circuit breaker can change the connection relation of the power system and adjust the power flow according to the demand when the system is running normally. In the event of a system failure, a circuit breaker can disconnect the fault current and cut off the electrical connection between the faulty device and the system. The operation information of circuit breaker is important reference data for daily operation and management of regional energy system.

Switching equipment in regional energy system has the characteristics of high air tightness and good sensitivity, and many parameters such as moisture content and contact temperature of gas-insulated composite electrical equipment are important parameters that need to be monitored online. The GIS online monitoring system mainly monitors two types of data, one is the density of sulfur hexafluoride gas, the other is the water content generated when a small amount of water enters the monitoring system as SF6. The partial discharge monitoring system the of GIS online monitoring system can monitor the defects such as conducting small particles and some alien substances, conducting current, internal air gap, and blocked grounding brought into the equipment during manufacturing, installation, and maintenance. GIS real-time system can determine the location of hidden danger by collecting different location information. By using sensors to collect temperature information inside the equipment, GIS high-speed optical fiber temperature measuring instrument can quickly and accurately obtain the temperature information inside the equipment. The data generated by the common primary equipment in the area energy system, such as isolation switch, grounding switch, capacitor, reactor, and transformer, plays an important role in the maintenance and fault diagnosis of the area energy system.

Environmental monitoring data are also of great significance to intelligent regional energy system data. The micrometeorological monitoring system of the intelligent regional energy system can regularly collect environmental data around the intelligent regional energy system, including air humidity, haze information, and other data.

2.2. Regional Energy System Data Classification. In order to facilitate the operation of data collection and processing, it is necessary to classify and process the data of the smart district energy system. Among them, the data volume of primary equipment is huge and diverse, and the related similar data can be roughly divided into five types for cluster analysis, namely, basic data, online monitoring data, operation data, test data, and accident data.

(1) Basic data: basic data refer to the ledger and design parameters of the primary equipment. The data are usually complete and accurate. Including the basic parameters of power equipment information, such as rating, power, size, manufacturer, and date of production. These data are permanently stored in the power equipment database and the other data are the flow data of the equipment.

(2) Online monitoring data: online monitoring data are the continuous or periodic automatic monitoring and detection of regional energy system equipment. It has the characteristics of high monitoring frequency and large data volume and can reflect the electrical, mechanical, and chemical characteristics of related equipment, such as the insulation oil chromatographic analysis of transformers and dielectric loss of capacitance bushing.

(3) Operating data: operating data refer to written or electronic records obtained after a device is inspected according to the specified check content and period during a device running. Operation data reflect the specific operation of electrical equipment, such as current, voltage, active power, reactive power, and circuit breaker operation times.

(4) Test data: test data can reflect the electrical, mechanical, chemical, and other properties of the equipment data and their specific value through the use of professional instrument test. Usually, equipment test data are obtained by the experiment after equipment power outage, such as DC resistance and insulation resistance, but it also includes in the case of equipment without power outage, far away from the equipment body data test, such as oil pressure value.

(5) Accident data: when a fault occurs in the regional energy system, the accident data mainly refers to the data of the related equipment when the short-circuit accident occurs, such as the effective value and peak value of the short-circuit current and the waveform of the short-circuit current.

3. Regional Energy Information Processing Based on HMM Model

Vector representation of regional energy information is required before regional energy information processing. The common method is to use vector group form composed of 0 and 1 to represent regional energy information, but there are problems of dimensional disaster and data sparsity, and association information in the same statement cannot be represented, which lead to word isolation [17].

In order to effectively vectorizing the associated relationship among fault information, word vector representation based on deep learning is adopted in this paper. The word vector is determined by calculating the distance between words, and the model is shown in Figure 1.

This method introduces word vector to construct probability model, and the formula is as follows:

$$
\begin{aligned}
p(s) &= p(\omega) \\
&= p(\omega_1, \omega_2, \ldots, \omega_T) \\
&= \prod_{t=1}^{T} p(\omega_t | \text{context}).
\end{aligned}
\tag{1}
$$

The words above are used to predict the words that may appear in the following paragraphs, and the hidden layer is used for nonlinear processing of the results. Finally, the corresponding probability value is obtained, and the maximum likelihood estimation is trained to obtain the word vector ω_t.

Variable regional energy information contains a large number of power system vocabularies but lacks the annotated corpus. Hidden Markov model (HMM) is used to process regional energy information. Each word in regional energy information has its own lexeme label. There are four kinds of morpheme labels in a sentence. B is the initial tag of the noun, M is the middle tag of the name, E is the end tag of the noun, and S is the single word tag. Each extracted information in the regional energy information record constitutes a corpus to be processed, and each word is labeled with lexeme. The word segmentation of regional energy information is summarized as A labeling problem. The word bit probability $\lambda = (\pi, A, B)$ of each word is obtained by HMM model, and the optimal labeling sequence is obtained by the Viterbi algorithm.

(1) State transition probability matrix: state probability of regional energy information word sequence.

Here, B, M, E, and S are the sequence of input words, and P is the probability of conversion between each state.

$$
A = \begin{bmatrix}
0 & P\left(\dfrac{M}{B}\right) & P\left(\dfrac{E}{B}\right) & 0 \\
0 & P\left(\dfrac{M}{M}\right) & P\left(\dfrac{E}{M}\right) & 0 \\
P\left(\dfrac{B}{E}\right) & 0 & 0 & P\left(\dfrac{S}{E}\right) \\
P\left(\dfrac{B}{S}\right) & 0 & 0 & P\left(\dfrac{S}{S}\right)
\end{bmatrix}.
\tag{2}
$$

(2) Observation probability matrix: the probability of obtaining each observation value according to the current state.
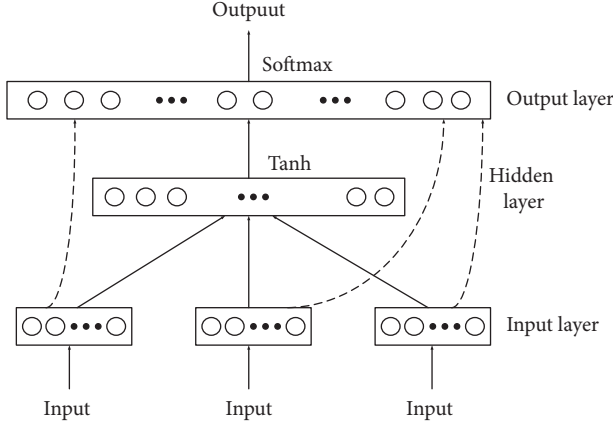
Figure 1: Word vector model.

Here, $P(o_n/Z)$ is the probability of the observed value, and o_n is the observed value.

$$B = \begin{bmatrix} P(o_1/B) & P(o_2/B) & \dots & P(o_nB) \\ P\left(\dfrac{o_1}{M}\right) & P\left(\dfrac{o_2}{M}\right) & \dots & P\left(\dfrac{o_n}{M}\right) \\ P\left(\dfrac{o_1}{E}\right) & P\left(\dfrac{o_2}{E}\right) & \dots & P\left(\dfrac{o_n}{E}\right) \\ P\left(\dfrac{o_1}{S}\right) & P\left(\dfrac{o_2}{S}\right) & \dots & P\left(\dfrac{o_n}{S}\right) \end{bmatrix}. \tag{3}$$

(3) State transition matrix: the probability of model transitions between states.

$$\begin{cases} S_t(i) = \max_{1 \le j \le N} \left[S_{t-1}(j) a_{ij} \right] b_i(O_t), \\ \phi_t(i) = \arg \max_{1 \le j \le N} \left[S_{t-1}(j) a_{ij} \right]. \end{cases} \tag{4}$$

(4) Optimal state sequence of energy information in the output region.

$$\begin{aligned} i_t^* &= \phi_{t+1}\left(i_{t+1}^*\right), \\ I^* &= \left(i_1^*, i_2^*, \dots, i_T^*\right). \end{aligned} \tag{5}$$

i_t^* represents the shortest path of regional energy information sequence, and I^* represents the optimal state sequence of regional energy information vocabulary.

4. Construction of Regional Energy Knowledge Map

Regional energy information has the characteristics of multidomain cross fusion and complexity. Knowledge graph is a structured semantic knowledge base, which extracts knowledge from text data in a structured way and forms the network knowledge structure of visual graph by connecting with each other [18]. Graph databases provide a unique perspective by focusing on the relational relationships between data. They

can examine proprietary data from different perspectives and even connect it to external data resources to further reveal the underlying relationships. Using the node parallel mechanism of graph database can improve computing performance and, at the same time, storing data and knowledge in graph database to build knowledge map, which can help answer the data knowledge and query questions raised by people in natural language communication. In addition, semantic network is used to organically connect all the original regional energy data from different sources for deep mining, which can find the neglected or difficult to detect connections between different data [19].

The Neo4j graph database uses a high-performance engine to realize the visualization function, which can map entity-relationship-attribute to knowledge graph. This paper analyzes and stores the regional energy information through Neo4j graph database, finds the internal correlation of the information, and realizes the connection of equipment information.

The construction process of regional energy knowledge map is as follows:

(1) To identify and define the professional field according to the characteristics of the electric power field to build the main power.

(2) The regional energy corpus is processed by word segmentation. The security risk information is extracted and processed by word segmentation method. Then, the word segmentation results are tagged with part-of-speech tagging, and then the entities and attributes of each part are identified by naming. Due to the specialty and particularity of the words in the electric power field, the professional dictionary of the electric power industry is imported into the database as an auxiliary word segmentation tool. If the extracted entities and attributes match the words in the dictionary, the entities and attributes will be determined. In the specific field of regional energy, the knowledge graph of safety hazards belongs to the closed graph, so the entity disambiguation step is not required.

(3) Knowledge processing: it mainly extracts the basic elements of the knowledge graph contained in the regional energy data, namely entities, relations and attributes, and the inherent hazards of hidden dangers of power equipment. Besides the entity-entity and entity-attribute relationships, it also needs to extract attribute-attribute relationships. Through the dependency syntax analysis of "subject-verb-object," "definite form complement" and other dependency relations among entities and attributes, and through the manual selection of labeling types of safety risk relations, the terms related to electric power safety regulations are divided in detail in order to form a corpus of safety risk relations. After the extraction of relationships, to avoid redundancy, semantic similarity calculation is used to screen redundant relationships [7] to improve processing performance.

(4) Knowledge fusion: use Neo4j graph database [20, 21] to integrate the processed data, import entity-attribute-relation triplet of regional energy and load corpus, create relations and attribute matching among entities, form knowledge map of regional energy, and realize global index. Figure 2 shows the example of generating knowledge atlas of substation security risks. Then, the basic elements of the knowledge map are integrated, and the regional energy knowledge map is shown in Figure 3.

(5) Knowledge update: with the continuous advancement of power system construction and the continuous operation of power system, the knowledge map of regional energy should be updated and amended to ensure the accuracy and effectiveness of retrieval based on the accumulation of fault information. The construction process of the regional energy knowledge map is shown in Figure 4.

The architecture design of regional energy knowledge atlas search engine is shown in Figure 5, which is mainly composed of data warehouse, service module, and application module. Data warehouse includes distributed management of elastic-block hidden metadata, Chinese word segmentation, index management, and logstash batch file import. After processing, the hidden data are submitted to the service module, where the elastic search service provides all sorting and read access, the data-offer service provides new data entries, and the security service provides permission control and security management. Then, the application modules are introduced, including hidden danger information input, retrieval display, and regional energy knowledge map display. The process of regional energy visualization system based on knowledge graph search engine is shown in Figure 2. The design steps of the regional energy knowledge map engine are as follows:

(1) Build Django background processing framework, which is mainly used for page request forwarding and processing, as well as creating elastic search engine service module and HMM-VA word segmentation service module.

(2) A front-end display platform based on vue.js is built, which is mainly used for the front-end display of background processing content, including regional energy retrieval information display, regional energy knowledge map generation, and statistical analysis functions.

(3) Environment module is deployed on Ali Cloud. On CentOS system, Docker is used to deploy Django background, HMM-VA word segmentation service module, elastic search, Vue front end, and other requests, configure relevant parameters, and use them jointly.

According to the above steps, by inputting regional energy information and adding specific key information fields, such as "66 kV," "substation," "rainy day," and other keywords, the 66 kV regional energy knowledge map can be

personalized and dynamically generated. The association search of the atlas can be used to analyze the types and causes of hidden dangers faced by all kinds of equipment, hazards after untreated results, possible treatment methods, violation of rules and regulations, prevention and control measures, etc.

5. Analysis of Experimental Examples

In this paper, regional energy data in recent three years are used as a data set to verify the effectiveness and practicability of the proposed method. The dataset contains information 1655 criteria. The hidden danger information is stored in the knowledge graph search engine by manual input and batch import. The proposed HMM word segmentation model is used for word segmentation processing of security hidden danger text. The validity of the proposed method is verified by the following four examples: comparison of regional energy data entity word segmentation methods; search engine information retrieval performance comparison; analysis of regional energy data cause knowledge atlas; and statistical analysis and prediction of regional energy data.

5.1. Comparison of Entity Segmentation Methods for Regional Energy Data. In this paper, the HMM model word segmentation steps, include corpus training, test set prediction, and word segmentation results. The power dictionary and regional energy data standard corpus of state grid corporation were used for pretraining, with a total of 150,000 words. The initial parameter information $\lambda=(\pi, A, B)$ was obtained by training the corpus in a supervised way instead of setting it as a fixed numerical parameter. The frequency of speech parts of energy data in each region was counted. The probability π can be calculated for the number of parts of speech and hidden speech of each region energy data and the corresponding words of the region energy data, and then the state transition probability matrix A and observation probability matrix B can be calculated.

Precision (P), recall (R), and F value, which are commonly used in entity word segmentation, were used as evaluation indexes to evaluate the word segmentation effect of regional energy data. Precision refers to the proportion of positive samples predicted by the model to positive samples in practice. The accuracy allows the classifier not to label negative sample errors as positive samples. Precision (P) is calculated in the following formula:

$$\text{precision} = \frac{TP}{TP + FP}, \tag{6}$$

where TP represents the number of entities in the test sample whose predicted category is positive and whose real category is positive. FP represents the number of entities whose prediction category is positive and real category is negative.

Recall rate refers to the proportion of positive samples that are predicted to be positive to positive samples that are actually positive. Recall rate can represent the ability of the
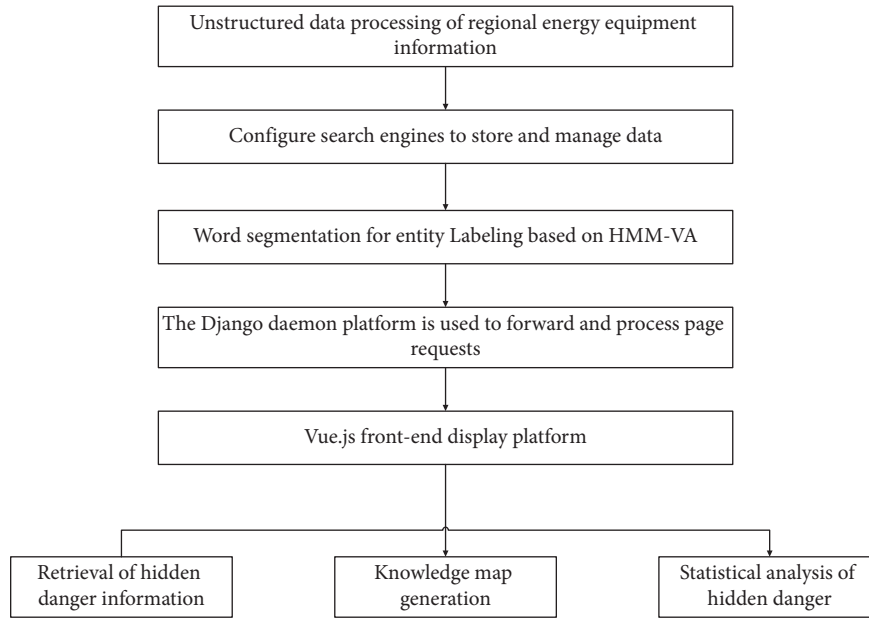
FIGURE 2: The process of regional energy visualization system based on knowledge graph search engine.
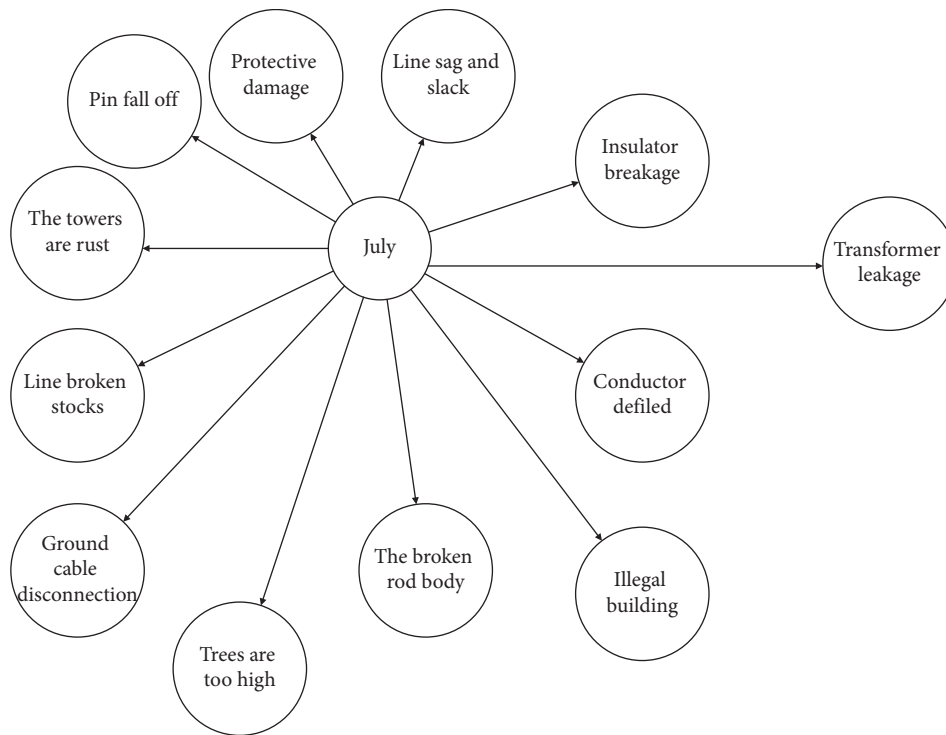

FIGURE 3: Knowledge Atlas of transformer Hidden dangers.

classifier to find all positive samples. The calculation of recall (R) is shown in the following formula:

$$recall = \frac{TP}{TP + FN}, \quad (7)$$

where FN is the number of entities whose prediction category is negative and real category is positive.

The harmonic mean of accuracy and recall is the F score. The specific calculation of F value is shown in the following formula:
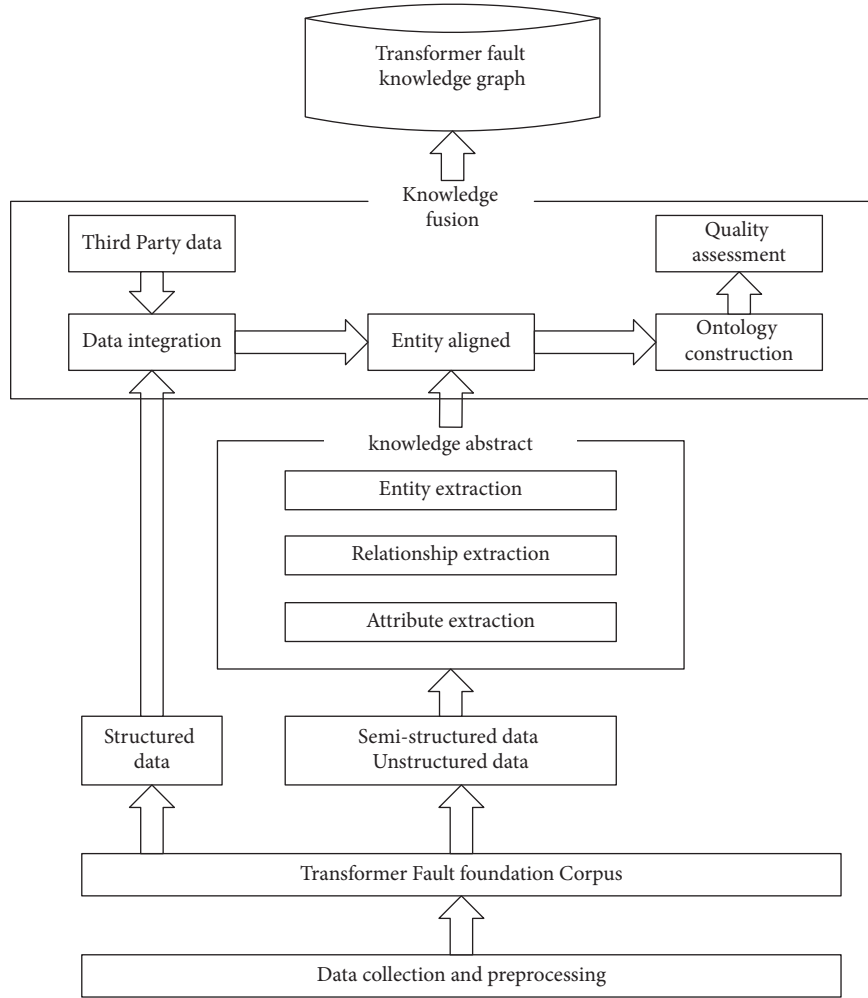
Figure 4: Knowledge graph construction process.

$$F = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}}. \tag{8}$$

The comparison of different named entity word segmentation models is considered to verify the word segmentation effect of the proposed model. Models involved in the experiment include BM matching model [22], N-Gram word segmentation model [23], Jieba model [24], and HMM-VA model proposed in this paper. The above models were all carried out on the same training set and test set. Table 1 shows the comparison results of the test set in different named entity recognition models.

As can be seen from the experimental results in Table 1, the average performance of the HMM-VA regional energy data segmentation model proposed in this paper is 85.93%. Higher than Jieba model 5.81%, N-gram model 10.8%, and BM model 15.42%, respectively, and better than the other three entity word segmentation models in terms of accuracy, recall rate, and F value.

5.2. Information Retrieval Performance Comparison of Search Engines.
The index performance of the system is measured by testing the index speed of standalone index [25], which is a

standalone search engine using the default configuration of elastic search, and distributed index, which is a 12-node distributed search engine designed and configured for the system. The test results for index performance are shown in Table 2.

As shown in Table 2, the elastic distributed search engine designed and configured in this paper is obviously superior to the standalone search engine in index efficiency indexes, such as average data rate, central processing unit (CPU) occupancy, memory occupancy (mem), and read/write rate (Io) and load rate, indicating that the proposed method can effectively improve the real-time performance of regional energy data index. It also meets the requirements of quick disposal efficiency in the investigation of hidden dangers in actual substations.

As shown in Table 3, for the search of the four test keywords, the average response time of the single machine is 1295.05 ms, and the average response time of the search proposed in this paper is 110.75 ms, indicating that the response time of the engine is significantly lower than that of the original single machine, which fully proves the advantages of the search engine in this paper. In addition, with the increasing of data volume, the processing advantage of the search engine for regional energy data is very significant.
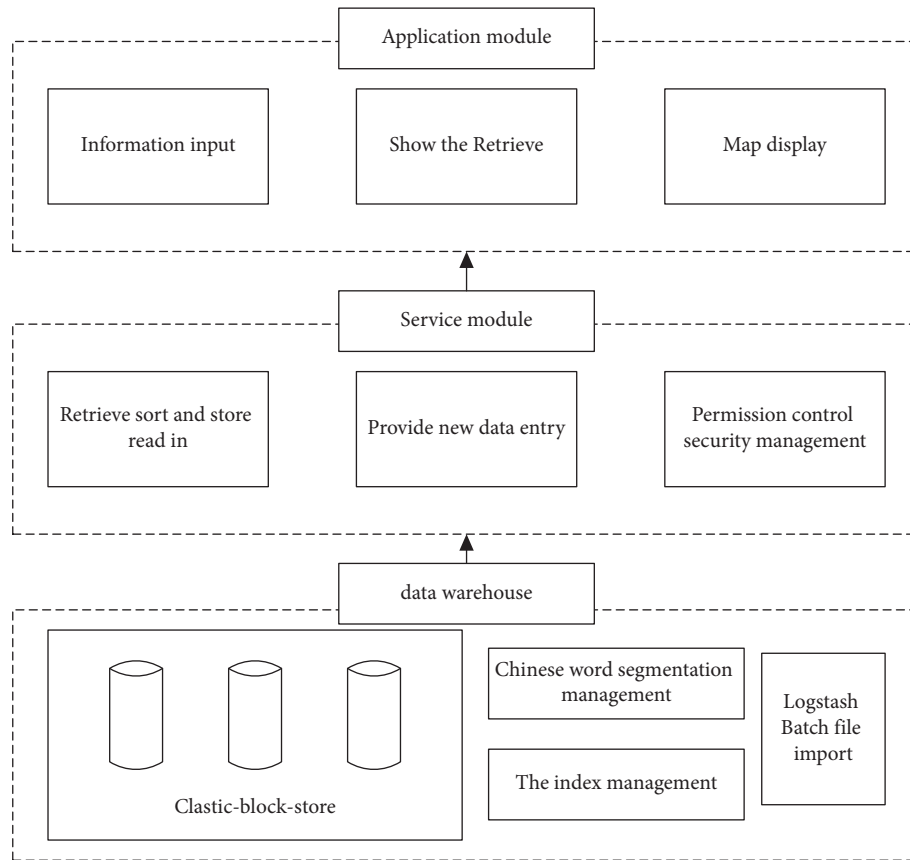
FIGURE 5: The architectural design of regional energy knowledge atlas search engine.

TABLE 1: Effect comparison of named entity recognition model.

Model	Precision (P) (%)	Recall (R) (%)	F score (F) (%)
BM matching	70.01	69.84	71.68
N-Gram word segmentation	76.77	74.27	74.35
Jieba model	81.52	80.31	78.54
HMM-VA	86.73	85.69	85.44

TABLE 2: Index test results.

	The average rate (thousand /SEC)	CPU (%)	Mem (%)	Io (%)	Load (%)
Single engine	6	57	63.6	90	43.1
The engine	39	14	16.5	41	18.7

TABLE 3: Search time test results.

Test key	Standalone search engine time (ms)	The system search response time (ms)
Substation	876	97
Rainy day	1438	161
Transformer	1389	84
Protection regulations	1479	101

6. Conclusion

This paper introduced technical knowledge map and flexible distributed search engine technology, this paper proposes a dynamic analysis method of regional energy data, through the research of regional energy data, distributed storage security hidden danger, to build knowledge map and analysis, has realized the knowledge map visualization display of search engine, and the safe hidden trouble efficiently retrieve and correlation analysis. It has good practical application value and promotion value. In the following research, more corpus features will be extracted in the relational extraction step to improve the accuracy of knowledge graph construction, so as to improve the analysis effect of regional energy data.

Data Availability

The dataset can be accessed upon request.

Conflicts of Interest

The author declares that there are no conflicts of interest.

References

[1] J. Lv, S. Zhang, H. Cheng et al., "A review of regional integrated energy system planning considering interconnection and interaction," *Proceedings of the csee*, vol. 41, no. 12, pp. 4001–4021, 2021.

[2] X. Zhang, W. Jing, and P. Wang, "Study on fault prediction method of high speed automata based on wavelet neural network," *Foreign electronic measurement technique*, vol. 39, no. 8, pp. 11–16, 2020.

[3] J. Cao, L. Chen, J. Qiu, and H. Wang, "Power grid defect text mining technology and its application based on semantic framework," *Power grid technology*, vol. 41, no. 02, pp. 637–643, 2017.

[4] G. Peng, Z. Zhou, and S. Tang, "Transformer fault prediction based on time series analysis and variable correction," *Electronic measurement technology*, vol. 41, no. 12, pp. 96–99, 2018.

[5] Y. Huang and X. Zhou, "Knowledge model for electric power big data based on ontology and semantic web," *CSEE Journal of Power and Energy Systems*, vol. 1, no. 1, pp. 19–27, 2015.

[6] Z. Wu, K. Bai, and L. Yang, "A review of text Mining for Electronic medical records," *Journal of Computer Research and Development*, vol. 58, no. 03, pp. 513–527, 2021.

[7] H. Tang, H. Huang, and Y. Tang, "Design of temperature and current measuring device for intelligent equipment," *Electric drive*, vol. 49, no. 05, pp. 93–96, 2019.

[8] Y. Wang, C. Peng, and Z. Wang, "Application of knowledge graph in power grid all-service unified data center," *Computer engineering and applications*, vol. 55, no. 15, pp. 104–109, 2019.

[9] J. Miao, J. Wang, and H. Zhang, "Research progress of uav fault diagnosis technology," *Chinese Journal of Scientific Instrument*, vol. 41, no. 09, pp. 56–69, 2020.

[10] Y. Fan, T. Ding, and Y. Sun, "Review and cogitation for worldwide spot market development to promote renewable energy accommodation," *Proceedings of the CSEE*, vol. 41, no. 5, pp. 1729–1752, 2021.

[11] Yi Ding, K. Xie, and Bo Pang, "Research on key issues of a unified national power market with Chinese characteristics (1): enlightenment, comparison and suggestions from foreign markets," *Power System Technology*, vol. 44, no. 7, pp. 2401–2410, 2020.

[12] Y. Song, M. Bao, and Yi Ding, "Summary of the key points of my country's power spot market construction under the new electricity reform and relevant suggestions," *Proceedings of the CSEE*, vol. 40, no. 10, pp. 3172–3187, 2020.

[13] Y. Xiao, X. Wang, and X. Wang, "Summary of research on electricity market facing high proportion of renewable energy," *Proceedings of the CSEE*, vol. 38, no. 3, pp. 663–674, 2018.

[14] C. Yue, C. Chen, Z. Liu, Z. Hu, and X. Wang, "Methodological function of CiteSpace knowledge graph," *Studies in Science of Science*, vol. 33, no. 2, pp. 242–253, 2015.

[15] W. Chen and H. Zhang, "Application of RF-LighTGBM algorithm in fan blade cracking fault prediction," *Electronic measurement technology*, vol. 43, no. 01, pp. 162–168, 2020.

[16] T. Liu, H. Chen, and G. Li, "Knowledge graph representation learning method for joint FOL rules," *Computer Engineering and Applications*, vol. 57, no. 4, pp. 100–107, 201.

[17] Li Mei-xuan, Y. Chun, and X. Jin, "Transformer fault diagnosis model based on combined prediction [J]," *Mathematical modeling and applications*, vol. 9, no. 4, pp. 49–56, 2020.

[18] X. Bian, L. Zhang, and Bo Zhou, "Based on knowledge atlas of the electric power market research both at home and abroad," *Journal of electrotechnics*, pp. 1–13, 2022.

[19] G. Liu, R. Dai, Yi Lu et al., "Power graph computing platform and its application in energy Internet," *Power Grid Technology*, vol. 45, no. 06, pp. 2051–2063, 201.

[20] Y. Ma and Z. Wu, "Modeling and analysis of power big data based on Neo4j," *New technology of electrical engineering and energy*, vol. 35, no. 02, pp. 24–30, 2016.

[21] H. Wang, Q. Zhang, and W. Cai, "Research on domain ontology storage based on Neo4j," *Application Research of Computers*, vol. 34, no. 08, pp. 2404–2407, 2017.

[22] Z. Yi, "Research on improved BM pattern matching algorithm based on network intrusion detection system," *Computer applications and software*, vol. 29, no. 11, pp. 193–195+207, 2012.

[23] F. Lizhou, G. Yang, and Xu Xue, "Based on N - "gramm bidirectional matching [J]," *Journal of Chinese word segmentation method of mathematical statistics and management*, vol. 33, pp. 633–643, 2020.

[24] P. Chen, X. Geng, M. Zou, and D. Tan, "Computer and Modernization," no. 3, pp. 77–81+92, 2020.

[25] D. Li and A. Madden, "Cascade embedding model for knowledge graph inference and retrieval," *Information Processing & Management*, vol. 56, no. 6, Article ID 102093, 2019.