Research on Intelligent Search Framework Technology Oriented to Monitoring of Power Grid Dispatching Equipment

Xiaohui Pan¹, Tan Chen² and Shengnan Liu*²

¹State Grid Jiangsu Electric Power Co., Ltd., Nanjing, Jiangsu, 210000, China
²Beijing Kedong Electric Power Control System Co., Ltd., Beijing, 100192, China

*Corresponding author’s e-mail: liushengnan1@sgepri.sgcc.com.cn

Abstract. The development of power grid dispatching and monitoring services puts forward new requirements for the digitization and intelligence of dispatching systems. This paper combined grid control data and business characteristics to study the intelligent search framework for grid dispatch monitoring, as well as key technologies such as thesaurus modeling, power feature word segmentation algorithm, multi-factor correlation ranking algorithm, and power knowledge graph association modeling, and developed a smart search application for the power grid that regulates and controls the cloud, to implement search for power grid equipment models, equipment operations, monitoring events, services, etc. It can carry out comprehensive, accurate, fast and intelligent retrieval of the monitoring operation data of mass dispatching equipment, which provides strong support for the management and decision-making of dispatching equipment monitoring business.

1. Introduction

With the construction of the power grid dispatching automation system, a dispatching automation technology system with clear levels and complete functions is gradually formed, which supports the safe and stable operation of the power grid. With the development of UHV (ultra-high voltage) and new energy access services, the scale of the power grid is expanding, the shape of the power grid is changing, and a wealth of regulation and operation data has been accumulated, which has promoted the trend of digital and intelligent development of power grid dispatch.

At the same time, the development of new IT technologies such as cloud-computing, big data, IoT, mobile internet and Artificial Intelligence also provides new ideas and means for intelligent dispatch transformation. The cloud platform provides the infrastructure for the digitization and intelligentization of the dispatch automation system. Combining with the operation status and development trend of the control system, advanced and mature IT technologies such as cloud computing and big data are introduced to build a physically distributed and logically centralized control cloud platform[1], to realize flexible allocation of resources, standardized management of data, and service-oriented packaging of business.

The regulation and control cloud uses cloud computing to virtualize resources such as servers, storage devices, and network devices, and establish computing, storage, and network resource pools to provide guarantee for the informationization of dispatching services. According to the structure of
general data objects for grid dispatching, with grid objects as the core of data standardization design, the model data cloud platform, operation data platform, real-time data cloud platform, and big data platform are designed to form standardized, integrated, and standardized control data. The system provides the basis for the digitalization of regulation and control business. At the same time, according to the design concept of Internet service and the characteristics of the regulation and control business, the business is split and packaged, and the service-centric application construction method is designed to form a clear business system, which improves the reuse of services. According to the Internet thinking such as rapid iterative and service sharing, applications are quickly built to meet business development.

The centralized monitoring of the control center faces the continuous expansion of the scale of the power grid, the increase in monitoring equipment, the number of monitoring information doubles, and the more complicated abnormal fault conditions. With the continuous access of data such as the control cloud model, operation, and management, the service system is gradually improved, and the business application is gradually rich. Faced with the problems of large data volume, complex types, diverse services, and difficult retrieval, the traditional portal website cannot meet the needs of business personnel to quickly obtain data.

Combined with the Internet search engine related technology, according to the characteristics of device monitoring business and regulating cloud data and services, this paper designed an intelligent search framework for device monitoring business, to realize the connection of various types of data such as monitoring device model, device operation, monitoring events, scheduling procedures. Data retrieval, comprehensive, fast, accurate and intelligent retrieval of data has become a new interactive method for regulating production operation managers and scheduling data.

2. Device monitoring search business analysis

2.1. Search scope
The main search scope of the search framework is to regulate the massive data stored in the cloud and construct a large number of services.

Based on the structured design standards and specifications, various business data of the control center such as models, operations, and grid data are collected in the control cloud. Among them, the model data includes basic grid models such as organizations, power containers, primary equipment, automation equipment, protection equipment, etc., and the operating data includes historical data of measurement, electricity, events, faults, alarms, plans, forecasts, loads, etc[2]. There are many types of these data and large amounts of data. The rapid search of business data such as account history and fault abnormalities can help the control personnel to make better use of the data and assist the operation and management of control.

According to service-oriented design specifications, a large number of services have been built in the control cloud, including public services, basic services, model services, data services, computing services, display services and interaction services. Each service has an independent operation unit, which integrates fixed business logic. The search for these services can assist the operation staff to quickly locate the relevant business services and meet the corresponding business needs.

| Search scope          | Organization | Power container | Measure ment data | Power data | Technical standard | Institutio nal norms | Public service | Basic service |
|-----------------------|--------------|-----------------|-------------------|------------|--------------------|--------------------|----------------|---------------|
| One-time equipment    | Automatic equipment | Event data     | Planning data     | Test report | Actual test report | Model service       | Data service   |               |

2
2.2. Requirements for intelligent search

Based on the massive amount of data and services, intelligent search needs to satisfy the increasingly complex power grid and deepen the challenges of dispatching and operating tasks. The dispatching and operating personnel have a comprehensive, accurate, fast, and intelligent access to data and services. It can improve the efficiency of operation management, use data to strengthen the support capacity of the power grid, and promote the development of intelligent and digital dispatch. Specific requirements are as follows:

- **Comprehensiveness of the search.** In response to complex incidents in the power grid, monitoring and operation personnel need to integrate various types of information for analysis and make decisions on dispatching operations. Therefore, the search framework should support the comprehensive search of various types of data such as models, operations, processes, and documents, and display all types of information in association, and push them to the monitoring and operation personnel for analysis and judgment.

- **Accuracy of the search.**
  Internet general search is more about showing the matched content to the user. The user judges the required information on the first two pages of the search result. However, the power grid is a precise physical system that requires high accuracy of information. How to accurately retrieve information based on search terms from a variety of data and services is the focus of an intelligent search framework.

- **Rapidness of the search.**
  Facing massive data such as models, operations, data, etc., it is necessary to build an efficient search model to quickly realize data search. According to the traditional standards of search engines, the search time is controlled within 2 seconds.

- **Intelligence of the search.**
  Relying on the topology of the power grid model, all auxiliary objects and related attributes in the model are mined to realize the intelligent drive from "single data retrieval" to multi-dimensional stereoscopic "panoramic information model".

- **Expandability of the search framework.**
  With the construction of control cloud, the continuous access of various types of data and the gradual construction of services, it is required that the intelligent search framework should have good scalability to quickly realize data access and service retrieval to meet the rapid development of regulation and control business, responding to the changing needs of search.

3. Research on key technologies of intelligent retrieval

The intelligent search framework studied in this paper mainly includes key technologies such as thesaurus modeling in the power field, power feature word segmentation algorithm, multi-factor relevance ranking, knowledge graph association modeling, etc., which provides the technical foundation for the construction of an intelligent monitoring framework for equipment monitoring.
3.1. Thesaurus construction technology in the field of electric power

The search engine's word segmentation algorithm, sorting algorithm, and knowledge graph construction are all based on text segmentation. English words are separated by spaces, and words are split according to spaces. Chinese words are continuous, and a Chinese lexicon needs to be built to extract Chinese words. As the field of exhibition industry, the power grid industry contains not only conventional Chinese words, but also the names of power grid equipment made up of professional vocabulary of the industry. Among them, there is no lack of place names and other conventional pre-corpus materials to form a comprehensive and complex name.

As shown above, this paper expands the technical terms of power system on the basis of the general vocabulary, such as "Xinghua Power Plant", "Jinling Substation", "Inrush Current", "Circuit Breaker", "Automation", etc. The words were modelled according to the hierarchy method. Using the hierarchical modeling method, it includes three levels of business sub-library, word label, and word items to ensure the standardization, scalability, universality, and expandability of the thesaurus management. After modeling, it can be reused in various power grid departments, especially at all levels of dispatch, and coordinated expansion.

There are two main ways to extract words: 1) Use the objects stored in the control cloud model table of existing power plants, substations, AC lines, DC lines, etc. to remove, de-duplicate, classify and store ambiguous words to the corresponding business Sub-library. 2) Use TextRank algorithm to extract keywords from unstructured documents such as scheduling procedures, stable quota procedures, technical standards, and specifications[3], and then classify them, store them in the corresponding business sub-library, and put those cannot be classified into technical standards sub-category.

Algorithm 1  TextRank keyword extraction algorithm.
Input: Input text data T such as dispatching regulation
Output: Keywords of Electric Power
1: Split the given text T by complete sentences
2: Form an array of sentences: T=[S1, S2, S3, ... Sn]
3: For S word segmentation, filter stop words and retain nouns
4: Get the word segmentation results of Si Keywords: Si=[ti,1,ti,2,ti,3,……ti,m]
5: Construct the resulting keyword graph $G = (V, E)$, where $V$ is the node set, composed of candidate words generated by $Si$, $E$ is the edge of the node, only when the number of the same words between the nodes is $M$, $M$ represents the window size.

6: Use the TextRank formula to iteratively calculate the weight of each node, until convergence.

7: Sort the nodes in reverse order of weights to obtain the highest ranked $T$ word.

8: Classify the acquired words and store them in the corresponding sub-library.

3.2. Power feature word segmentation algorithm

Generally speaking, the word segmentation algorithm [4] is involved in two cases. One is to segment the search term’s advancedness, and then to match in the index. The other is to build the search index after segmenting the data for models, operations, documents and other data. Based on the thesaurus of power field and the accuracy requirements of power word segmentation, this paper adopts the maximum forward matching word segmentation method to implement the word segmentation algorithm based on power characteristics.

The overall process of power word segmentation is: 1) Filter the empty words out of the input words first into useless mood words, adjectives, etc. 2) Use electric power thesaurus and maximum forward matching algorithm to split the sentence string to obtain the word segmentation result. 3) Form a participle list and return. For example: for the search term "What is the power generation capacity of #6 unit in A power plant", where "A power plant, unit, power generation capacity" are special words in the power vocabulary, and the final split result is an array [A power plant, #6, unit, power generation capacity]. The overall process is shown below:

![Diagram](https://via.placeholder.com/150)

**Figure 3. Power feature segmentation process.**
In this process, the core step is the maximum forward matching algorithm. This algorithm has high operation efficiency and is suitable for word segmentation algorithms that need to be called frequently. The time complexity is O(n). The specific algorithm flow is as follows:

Algorithm 2 Maximum forward matching algorithm.
Input: Text or search words with word segmentation, power thesaurus
Output: Array of word segmentation results
1: Calculate the character length of the input string
2: Record and match start position
3: while The starting position of the record is less than the length of the string do
4: Record the word with the maximum length in the forward direction
5: if the word already matches a word in the thesaurus then
6: Output the word and move the pointer to the next digit
7: else
8: Split according to a single word, output a single word and point the pointer to the next bit
9: return device name division group

3.3. Multi-factor correlation ranking

TF-IDF is a weighted sorting algorithm commonly used in information retrieval. It is used to evaluate the matching degree of a word to a file or data [5], and the matching score can be calculated. The TF-IDF core has two points: 1) If a word appears more frequently in the file, the more important the word, the higher the score, that is TF; 2) If a word appears in many files, the word's ability to distinguish between files is low, and the score is also low, that is, IDF.

$$\text{TF-IDF} = \text{TF} \times \text{IDF}$$

TF calculation formula: For the word $t_i$ in a file $d_j$, the word frequency $t_f$ of $t_i$ can be expressed as:

$$t_f_{i,j} = \frac{n_{i,j}}{\sum n_{k,j}}$$

Where, $n_{i,j}$ is the number of occurrences of the word $t_i$ in the file $d_j$, and the denominator is the sum of the number of occurrences of all words in the file $d_j$.

The IDF value of a specific word can be divided by the total number of files divided by the number of files containing the word, and then obtained by taking the logarithm of the obtained quotient:

Where $|D|$ is the total number of all files in the corpus, and the denominator is the total number of all files containing the word $t_i$.

$$\text{idf}_i = \log \frac{|D|}{|\{j: t_i \in d_j\}|}$$

The sorting results obtained by the TF-IDF algorithm are relatively accurate for the search and ranking of webpages. Basically, webpages are all of limited text length, and contain fewer keywords. The use of TF-IDF to distinguish webpages usually has better results. However, according to the regulation of cloud data search in this paper, the characteristics of the data are relatively different, mainly reflected in three aspects: 1) There is a lot of small data, that is, data with less text, such as the model parameters of a power plant. The text information contained in it mainly includes the name of the power plant, the voltage level, the power grid, and the scope of regulation. If we simply follow TF-IDF algorithm will result in a lower IF score, the model parameter information itself is more important, and the ranking will be lower. 2) There is a part of big data, that is, data with long texts and a large number of keywords, mainly unstructured files. The unstructured files of the power grid are mainly files such as dispatch procedures, technical standards, and technical specifications. The file has
the characteristics of a long space and a large amount of information. For example, in the dispatch procedure, there will be many power characteristics words such as maintenance, plan, fault, unit, etc., and these words will appear frequently, which will lead to a higher TF value. Regulating the amount of data in the cloud search engine will result in a lower IDF value, which will often cause documents to overwrite other more professional information. 3) In addition to the matching score obtained by TF-IDF, the factors that affect the grid search include the matching degree of the data type and the search term, the timeliness of the data, the user’s specialty, or the grid.

Therefore, based on the TF-IDF score, this paper introduces an improved multi-factor correlation ranking algorithm, including TF-IDF score, data type score T, timeliness score M, and affiliated grid score D, through the weighting coefficient to form a unified correlation calculation formula, returning more concerned information to the scheduling user.

\[
\text{score} = (\text{TF-IDF}) + t \cdot T + m \cdot M + d \cdot D
\]  

(4)

In the above formula, T is the data type score and t is the coefficient. For example, when the user search term is a power plant name, the model data type score is higher. M is the timeliness score, m is the calculation coefficient, when searching historical data, the latest data score is enlarged. D is the user's professional and grid score. For example, when a user searches for a grid fault, the fault in the grid where the user is located will be ranked first.

3.4. Knowledge graph association modeling

The key technologies above mainly solve the problems of comprehensive search, accurate search and fast search. New human-computer interaction methods such as natural language analysis and retrieval provided by existing search engines in the Internet field, and Doren interactive question answering provided by smart speakers are all supported by knowledge graphs. Therefore, this paper takes advantage of the more standardized and structured design of the control cloud and the continuous improvement of data quality, introduces knowledge graph association modeling, builds a knowledge graph in the field of power grid dispatching [6], and provides a "question-and-answer" search experience. The main modeling and usage methods are as follows:
What is the total installed capacity of wind power plants in XX area?

Query using knowledge graph

As shown in the figure above, based on a structured grid control model, intelligent search uses the modeling theory of ontology to analyze the association between various power object entities (power plants, generators, etc.), and establish the topological relationship between power entity objects and object attributes, objects events, building a knowledge graph model of power ontology. First determine the ontology object (such as thermal power plant), and then use this ontology object as the center to determine the scope and topological relationship between the central ontology and other subsidiary ontology (such as power plant power, power plant output), and gradually form the ontology model with center ontology as the core, and other related ontology as the branches. Finally, the relevant underlying attribute data (such as coal consumption and power generation amount) is drilled to form a complete power knowledge map, which provides the basis for intelligent reasoning and question and answer search. For example, ask "What is the total installed capacity of wind power plants in XX area?" Such questions can be directly calculated and returned to the user using the knowledge graph.
4. Design and application of intelligent retrieval framework

In this paper, based on the research results of key technologies such as thesaurus in the field of power grid, power feature word segmentation, multi-factor correlation ranking, control knowledge map and other key technologies, combined with the control cloud service bus, a set of comprehensive, accurate, fast, intelligent and expandable overall framework was designed.

As shown above, the intelligent search framework for control cloud mainly includes three parts: data sources, basic services, and business services. 1) Data source: Mainly use the model data platform, operation data platform, and big data platform of the control cloud to obtain structured data, and use the file service to obtain unstructured power grid information. 2) Basic services: It includes data collector, thesaurus of power grid, knowledge graph, data index, ranking service and word segmentation service. Among them, the data collector provides a configuration management tool, which can dynamically expand the data source, realize rapid access to various types of data such as models and operations, and meet the requirements of comprehensiveness and scalability of intelligent search. Index storage enables all kinds of collected data to be stored according to the inverted index, which can increase the speed of data retrieval and meet the requirements of fast search. The four services of the thesaurus, knowledge graph, sorting algorithm, and word segmentation algorithm in the field of power grid have realized the accuracy and intelligence of the search.

3) Business service: Use the regulation and control cloud service bus, use basic services to realize model data services, power grid event services, power grid data services, intelligent inference services, etc., and to realize data search. Then, the extensibility of servitization is used to provide extensibility for the change of the following business. At the same time, access to public services, data services, computing services and other control cloud standard services in control cloud services to achieve service search.

As shown in the figure above, it is the application effect of the intelligent search framework, in order to evaluate and verify the effect of intelligent search. According to the requirements of intelligent search, the search terms for verification are selected to define two evaluation indicators: search speed and accuracy of the top ten search results. The verification results are as follows:

| Search term          | Search speed (s) | Accuracy rate (%) |
|----------------------|------------------|-------------------|
| Jinling Substation   | 1.2              | 100               |
After verification in this paper, based on various structured, semi-structured, and unstructured data resources in the field of power grid regulation, a power regulation intelligent search engine "regulation Baidu" was constructed to achieve comprehensive, accurate, fast, and intelligent search of power regulation data.

5. Conclusion
Research on grid intelligent search technology for power grid dispatching equipment monitoring will quickly and comprehensively retrieve all kinds of structured and unstructured data such as model, operation, and documents of the control cloud, and support the search and control of cloud services to dispatch the power grid. The professional terminology in the field of equipment monitoring is combed to form a power vocabulary, which improves the accuracy of data retrieval. The scheduling knowledge graph formed according to the model topological relationship provides the basis for intelligent retrieval of data, and provides technology for subsequent application scenarios such as reasoning and question answering. Guarantee. The design of the framework better opens up the standardized data and services of the control cloud, provides a convenient and unified data access entrance for dispatching, monitoring, operation and management staff, improves the data application level of the control system, and allows employees to freely and conveniently retrieve data. At the same time, it also provides the basis for the subsequent ubiquitous IoT to open business data services to external users, and can explore the requirements of ubiquitous IoT applications in the field of regulation and control.

References
[1] Xu, H.Q. (2017) The architecture for dispatching and control cloud and its application prospect. J. Power System Technology, 41(10):3104-3111(in Chinese).
[2] Xu, H.Q. (2018) Structure design and application of general data object for power dispatch based on control cloud. J. Power System Technology, 42(07):2248-2254(in Chinese).
[3] Xu, L. (2019) Text keyword extraction method based on weighted TextRank. J. Computer Science, 46(S1):142-145(in Chinese).
[4] Cheng, Y.S., Shi, Y.T. (2018) Chinese word segmentation method for professional field. J. Computer Engineering and Applications, 54(17):30-34+109(in Chinese).
[5] Chen, Z.F., Pang, F., Zhang, T., Tian, Y. (2019) Research and Application of Intelligent Understanding and Matching Method of Power Equipment Information. J. Computer Knowledge and Technology, 15(15):250-251+254(in Chinese).
[6] Guan, S.P., Jin, X.L., Jia, Y.T., Wang, Y.Z., Cheng, X.Q. (2018) Research progress of knowledge reasoning for knowledge graph. J. Journal of Software, 29(10):2966-2994(in Chinese).
[7] Zhao, H.B., Huan, Y.L., Chen, S. (2003) Professional intelligent search engine based on power plant operation database. J. Electric Power Automation Equipment, 23(8):25-28(in Chinese).
[8] Shang, X.W., Zhao, L., Fan, Z.L., Ye, F., Fan, G.M., Guo, L.X. (2018) Architecture and key technology of wide area data bus system based on dispatching data network. J. Automation of Electric Power Systems, 42(11):109-114(in Chinese).