Research on Construction Technology of Multi Heterogeneous Data Resource Graph of Power Grid Corporation

Liying Wang¹*, Shi Tian¹, Liang Shen², Zhihua Cheng², Ming Zeng¹ and Yanbin Li¹

¹North China Electric Power University, Beijing, 102206, China
²Big Data Center, State Grid Corporation of China, Xicheng District, Beijing, 100031, China

*Corresponding author e-mail: 595910434@qq.com

Abstract. The development of new-generation artificial intelligence technologies such as deep learning, knowledge graphing, and enhanced learning is driving the "Internet + Smart Energy" into a new era of "Internet of Energy." The data structure, type, and format of the data between the main body of the energy and power system, various departments of the enterprise, and various types of users are all constructed at different times using different standards. The data is huge and complex. Establishing a company-level data center for power grid companies to resolve the standardized management and fast and reliable exchange of heterogeneous data under massive data is an opportunity and a challenge for the company to build an energy Internet. As the core foundation of integrating heterogeneous data resources of power grid, data resource graph can provide effective support for the management of heterogeneous data of power grid enterprises under the energy Internet. From the multi-dimensional perspective of data modeling, resource management, data navigation, learning cognition, database, this paper focuses on the diversified and heterogeneous data resources of power grid companies, and explores from the specific technical aspects of data ontology modeling, data naming and identification, data relationship mining, data fusion and exchange technology, aiming to provide technical support and reference for power grid companies to build data resource graph.

Key words: Knowledge graph; Data resource graph; Data center; Energy Internet.

1. Introduction

With the growth of data resources, the improvement of computing power and the enhancement of machine learning algorithms, artificial intelligence has ushered in the third wave of development after several ups and downs. Machine learning represented by deep learning and knowledge engineering represented by knowledge graph have become the core technologies to realize disruptive changes in many fields. In computer vision, image processing, speech recognition, natural language processing, machine learning, machine learning represented by deep learning has a very broad and basic application. However, bottlenecks such as the need for large quantities of annotated data, the uninterpretability of forecast results, and the difficulty in combining human intentions and domain knowledge are also
increasingly prominent. At the same time, knowledge graph provides a powerful supplement to deep learning, showing great power in semantic search, natural language understanding, human-computer interaction dialogue, logical reasoning and other aspects, and has become the basis of the current knowledge-driven intelligent application. Therefore, the field of energy and power systems has also begun to pay attention to the important role of knowledge graphing. Deep learning, knowledge graphing, enhanced learning and other new generation of artificial intelligence technologies will become the core driving force for "Internet + smart energy" to enter the new era of "energy Internet". At the same time, the differences in data structure, type and format between the main body of each link of the energy and power system, various departments of the enterprise and various users, as well as the huge scale of data, complex relationships and non-uniform standards have also brought challenges to the construction of the company's energy Internet.

In this context, many scholars have conducted in-depth research on how to solve the problems of standardized management and fast and reliable exchange of heterogeneous data under the massive data of grid companies. Literature [1] converts multi-source heterogeneous data from different sources into RDF format data through the RDF interface, and combines SWRL rules to solve the problem of multi-source heterogeneous data fusion in power transmission and transformation systems; Literature [2] uses XML technology to achieve access between different databases of the power grid and real-time data sharing; Literature [3] proposed a mass data storage method for the characteristics of structured data and unstructured data, in order to solve the storage and access problems of large power data; Reference [4] aimed at the "data island" phenomenon in the development of smart grids, and proposed the concept of a real-time database to realize the rapid storage and access of big data in electricity.

However, combined with the actual situation of power grid companies, the above research results still have problems such as low operation efficiency, slow relationship search, unable to meet the requirements of data visualization, and real-time requirements of data query. In view of this, this article introduces the concept of data resource graph in the field of energy and power systems, which aims to solve the problems of data diversification and "data islands" mentioned above. First, based on the basic concept of knowledge graph, the definition of data resource graph is summarized; second, the connotation of multi-dimensional heterogeneous data resource graph of power grid is analyzed from a multi-dimensional perspective; Thirdly, the technical framework of the data resource graph of the power grid company is constructed, and the key technologies involved are summarized; Finally, the application prospects of data resource graph in the context of Energy Internet are analyzed and discussed, with a view to providing some references for the application of data resource graph in the construction of Energy Internet.

2. Connotation, classification and technical framework of data resource graph for power grid companies

2.1. Relevant Overview of Data Resource Graph

Knowledge graph is a branch of the field of artificial intelligence, the most important way of knowledge representation in the era of big data. By combining the theoretical methods of applied mathematics, graphics, information visualization technology, information science and other disciplines with metrological citation analysis, co-occurrence analysis and other methods, and using the visual graph to vividly display the core structure, development history, frontier fields, and overall knowledge framework of the discipline. The knowledge graph can reveal the dynamic development law of the knowledge field through data mining, information processing, knowledge measurement and graphic drawing, and provide a practical and valuable reference for subject research. So far, its practical applications have gradually expanded in developed countries and achieved good results. The applications in different fields are shown in the following table.
As an important existing form of resources, data resources are stored in non-paper media carriers such as optical discs, hard disks, flash memories, and reproduced through network communications, computers or mobile terminals. Data resources are divided into dynamic and static types, which are specifically presented in various forms such as text, video, audio, and animation. Therefore, in order to achieve unified collection of multi-source heterogeneous data and improve the ability of data search and search, it is necessary to build a data resource graph.

The concept of data resource graph has evolved from knowledge graph, but at present, there is no clear definition of data resource graph in academia. This paper considers that data resource graph is a modern theory to achieve the purpose of multi-type data fusion and exchange. It refers to the use of the new generation of artificial intelligence technologies, such as deep learning, knowledge graph and big data, to visualize the relationship among data entities, data attributes and data values, etc. By organizing multi-source heterogeneous data in a more orderly and organic way, data resource graphs aims to provide users with more intelligent access interfaces, so that users can access the data they need more quickly and accurately, and make certain data mining, analysis and decision-making. Therefore, the core of data resource graph is the fusion of knowledge graph and data. By studying the construction of multivariate heterogeneous data resource database, data ontology modeling, data naming and identification, data relationship mining and data fusion and exchange, it has the ability to conduct large-scale association analysis of data of different structures, types and formats, providing effective support for the management of multivariate heterogeneous data in enterprises.

2.2. Connotation of data resource graph of power grid companies

The data resource graph is a knowledge graph based on a graph database. A graph database is a type of NoSQL (Not Only SQL) database that stores and queries data in a "graph" data structure. Its data model is mainly represented by entities (nodes) and relationships (edges), and can also handle key-value pairs. Its advantage is to quickly solve complex relationship problems.

The basic elements of a data resource graph are entities, attributes, and relationships. The entity refers to the object managed by the company. It is an object that exists objectively and can be distinguished from each other. It is represented by a unique ID value, and the entity characteristic information is recorded in the form of attribute-attribute value. The relationship information between different entities is recorded by the relationship. Attribute refers to a certain characteristic of an entity. An entity can be described by several attributes. It has two forms: object and attribute value. The object is the type and function of the entity. The attribute value is the value of the attribute specified by the entity. Relationship refers to the relationship between different sets of entities. There are one-to-one, one-to-many, and many-to-many relationships between entities. With reference to the knowledge graph, the relationship between entities, attributes, and relationships in the data resource graph can be expressed in the form of triples. The basic forms of triples include (entity 1-relation-entity 2) and (entity-attribute-attribute value).
This paper studies the data resource graphs of power grid companies in the field of energy and power systems from a multi-dimensional perspective such as data modeling, resource management, data navigation, learning cognition, and databases. From the perspective of data modeling, the data resource graphs of power grid companies is a method for modeling data entities. From the perspective of resource management, power grid corporation data resource graph is a way to provide one-stop data resource management service in the form of "graph" and semantically organize power grid corporation data resources and their relationships. From the perspective of data navigation, the grid company data resource graph can generate personalized learning paths oriented to learning goals with the support of big data, artificial intelligence and other technologies. From the perspective of learning cognition, state information can be mastered by superimposing the knowledge of data users based on the data resource graph of power grid companies, which can form the cognitive schema of data users. From the perspective of database, grid corporation data resource graph is a kind of structured semantic database storing grid data domain knowledge in a way that can be "understood" by computer.

The grid data resource graphs can be divided into two categories: one is the static data resource graphs, which takes the data entities involved in the process of data application as nodes and the logical relationship between data entities as edges to form a semantic network. The data entity here can refer not only to the data types involved in a certain business type of the power grid company (electricity and electricity statistics, equipment failure statistics, raw material inventory data, etc.), but also to electricity consumption, power grid investment, installed capacity of equipment, line loss, power supply reliability, etc.; The second is the dynamic data resource graph, which is a multi-relational graph composed of logical relations (sequence, causality, inversion, condition, upper and lower positions, composition, etc.), represented by activities such as power grid operation and maintenance. FIG. 1 and FIG. 2 are local graphs of grid static data resource graph and local graphs of grid dynamic data resource graph.

Figure 1. Local graphs of grid static data resource graph.
2.3. Data resource graph construction process of power grid company

2.3.1. Diversified and heterogeneous data resource types of power grid companies. Power grid companies have a large number of data resources for scientific research and learning. These multivariate and heterogeneous data can be divided into structured, semi-structured and unstructured data according to storage types.

Structured data is the main data source of data resource graph and dominated by relational data tables which mainly refers to the operation and inspection, marketing, materials and other databases of relevant departments of power grid companies. By establishing the graph relationship between the concepts in the database and the ontology in the data resource graph and rule-based reasoning, the data entities, attributes and their relations are extracted automatically from the database.

Semi-structured data mainly refers to the dispatching operation index and some webpage data obtained from the measurement and collection data of uhv substation. It is a kind of sequential regular data. The data collected from power grid equipment mainly include basic information of power grid and unit, power generation and load data of power grid, current and voltage data of equipment, declared transaction data of various transactions, AGC and various indicator data of unit, etc. It uses Hadoop big data technology to ensure efficient data transmission and storage through column storage and sparse matrix. The index information of semi-structured data, such as frequency, voltage, electric quantity balance index, power flow index of important constrained section, overhaul plan index, is obtained by day.

Unstructured data mainly refers to some text image data. The traditional word packet model segmentation method relies too much on dictionaries. For example, the text "general alarm of 115 line switch accident of xixing transformer" is normally understood as "xixing transformer", "115 line", "switch", "accident" and "alarm", which is easily divided into "xixing transformer", "change", "115 line", "switch", "accident", "total" and "alarm". For such data, existing data resource graph knowledge can be used to construct the training set by means of remote supervision. The extractor can be learned by deep learning and the extractor can be used to extract the data in the text.
2.3.2. **Data resource graph construction process.** Combined with the general domain knowledge graph construction process, the grid company data resource graph construction process is shown in the following figure. First, the basic data of relevant departments of the power grid and the operation data and text images obtained from the collection and measurement are selected and divided into structured data, semi-structured data and unstructured data. Secondly, data entities, attributes and relationship systems are constructed through data ontology construction, and data entities, data attributes and data relationships are extracted from them. Finally, for heterogeneous data with numerous sources and channels, data entity links are made through data naming and identification, data relationship mining and data fusion and exchange, so as to construct the data resource graph.

3. **The key technology of constructing multi-component and heterogeneous data resource graph of power grid corporation**

3.1. **Technical framework of grid company data resource graph**

In the study of general knowledge graphs, the technical framework of knowledge graphs is proposed in literature [5], which can be divided into three levels: information extraction layer, knowledge fusion layer and knowledge processing layer. According to literature [6], knowledge graph involves four technologies: knowledge extraction, knowledge representation, knowledge fusion and knowledge reasoning. According to literature [7], key technologies of knowledge graph construction include entity relationship recognition technology, knowledge fusion technology, entity link technology, etc.

By referring to the research results of the above knowledge graph, this paper constructs the technical framework of the grid company's data resource graph from four aspects: data ontology modeling, data naming and identification, data relationship mining and data fusion and exchange. Data ontology modeling is a framework to describe the data system using ontology construction method and to constitute the pattern of data resource graph. Data naming recognition is to identify data objects in data resources through information extraction technology. On the basis of data naming and recognition, data relation mining is to further judge the semantic association between data. Data fusion exchange is a
process of data information fusion from multiple data sources, eliminating ambiguous information in the process of information extraction.

Figure 4. Technical framework of data resource graphs.

3.2. Data ontology modeling technology
Data ontology is the framework and abstract model of grid corporation's data resource graphs, and can greatly assist the construction process of grid corporation's data resource graphs as the bottom pattern structure of the graphs. Data ontology modeling is mainly about the acquisition of data entities and the acquisition of relationships between data entities, including data entity types, attributes of data and semantic relationship types [8]. At present, there are three main methods to construct the data ontology of power grid companies: manual method, automatic method and semi-automatic method.

Manual method refers to manually constructing data ontology with the assistance and guidance of a large number of subject experts. Currently, the commonly used construction method is the seven-step method developed by Stanford university, which includes seven steps: determining the domain and scope of ontology, investigating the possibility of ontology reuse, listing concepts in ontology, defining the relationship between concepts, defining the attributes of concepts, defining the facets of attributes, and creating instances. In the process of ontology construction, we can also refer to or reuse the ontology patterns of existing databases, such as WordNet in English and HowNet in Chinese. Editing tools developed to facilitate ontology construction mainly include Protege, WebOnto, KAON, etc. Although these tools can greatly reduce the workload and technical threshold of manual ontology construction, they still need to edit data types, data attributes, data relationships one by one, which is inefficient, time-consuming and laborious.

Automatic method refers to the use of data acquisition technology, machine learning technology and statistical analysis technology to automatically extract the required data ontology from database resources, thus reducing the cost of manual construction [9]. There are three main methods for automatically constructing data ontology: automatic construction method based on data name, automatic construction method based on database and automatic construction method based on ontology learning [10]. These methods are only applicable to the domain with simple ontology structure, so their practicability and accuracy are not high for power grid companies.

Semi-automatic method, which is between manual method and automatic method, guides the construction of ontology model by using technical means and expert knowledge in related fields, which can not only reduce the expert labor force but also improve the accuracy of ontology construction.

3.3. Data naming and recognition technology
Data is the most basic element of grid corporation's data resource graph, while data naming and recognition is the most basic and critical step in the process of grid corporation's data resource graph construction. At present, the methods used in data naming and recognition technology include database and rule based method, statistical machine learning method and deep learning method.

(1) Method based on database and rules
This method establishes the corresponding domain database or data rule template by the power domain experts and industry workers by hand, and then uses the regular expression to match. In the process of template formulation, data structure, data type and data format are mainly considered. Although the identification accuracy of this method is high, it is only suitable for simple identification systems because of the problems of time consumption and poor domain migration in database construction or data rule template formulation.

(2) Method based on statistical machine learning

Methods based on statistical machine learning include hidden markov model, maximum entropy model, conditional random field model, etc. Among them, the hidden markov model has achieved good results in data naming and recognition. This method simplifies the task of data naming and recognition to the problem of sequence labeling, that is, given an observation sequence, a joint probability is solved to optimize the labeling sequence [11]. The maximum entropy model solves the characteristic function through the maximum principle of entropy [40]. Conditional random field model applies conditional probability principle to deal with sequence labeling problem [12]. The above method requires a large number of manually extracted features to be fed into the model to mark the data. Although it can achieve good results, the process of manually extracted features increases the difficulty of model construction.

(3) Method based on deep learning

Statistical machine learning methods rely heavily on feature engineering, and the biggest advantage of deep learning lies in the automatic learning of domain data features. Therefore, in recent years, deep learning has been gradually applied to data naming and recognition tasks. In 2011, Collobert earlier proposed a data naming and recognition method based on neural network model combined with window method and syntactic analysis [13], whose recognition effect and performance exceeded traditional statistical machine learning methods.

In order to further improve the accuracy of identification results, deep learning and statistical machine learning began to be combined in the research of data naming and recognition after 2015. For example, literature [14] et al. proposed the bilstm-crf model after combining the two-way LSTM network with the CRF model. The advantages of this model are high accuracy and low dependence on data volume. Subsequently, literature [15] further integrated CNN network and proposed an "end-to-end" named entity recognition model -- bilstm-cnn-crf model. CNN network is mainly used to learn the word vector features of data, and its recognition effect has been greatly improved. But overall, deep learning-based methods are still in the exploratory stage.

3.4. Data relationship mining technology

Data relationship mining can determine whether the identified data has a specified semantic relationship type. At present, the mainstream methods of data relationship mining technology can be divided into the following two kinds: rule based data relationship mining technology and machine learning-based data relationship mining technology.

For the structured types of data in the data center, the rule-based data relationship mining method is mainly used to analyze the table structure of data storage and obtain the relationship through the primary and foreign keys of the table and other information. Rules-based relationship extraction is to use the pre-established rules for matching, and the information that conforms to the rules is the extracted relationship. By analyzing the table structure of the data store, the primary foreign key information of the table is set as rules through which relational extraction of the data table can be performed. The advantage of rule-based relationship extraction is high accuracy, while the disadvantages are high labor cost and lack of information coverage.

For semi-structured data and unstructured type data in data, the law of data relationship cannot be clarified. Therefore, it is necessary to study the relationship extraction based on machine learning. Based on supervised, unsupervised and semi-supervised models, the relationship between data entities can be obtained through data analysis, joint reasoning and other methods.

The supervised model learns from the training data set to predict the relationship types of the test data, which are mainly divided into two categories. One is the method based on the feature vector; The
second method is based on the kernel acquaintance function. Supervised extraction can achieve good accuracy and recall rate on the same type of data set, but it cannot identify the relationships not contained in the corpus, so it is not suitable for processing large-scale corpus in open domain.

The unsupervised method does not need the corpus support processed in advance and can automatically extract the entity relations contained in the text, which can be divided into three categories. The first is the acquisition of named entity and its context. The second is named entity clustering; The third is to choose the core vocabulary to label all kinds of semantic relations. Unsupervised extraction has good portability and is suitable for processing large-scale open domain data, but the accuracy and recall rate are relatively low.

Semi-supervision method takes pre-defined relationship patterns and relationship instances as seeds, and finds new relationship templates and instances through machine learning, which can be roughly divided into two processes. First, look for statements that contain two entities with known relationships; Second, these statements are classified as training sets. Combined the advantages of supervised and unsupervised, the semi-supervised method does not need too much training corpus and can find the relationship examples not included in the seed set.

The above researches regard data naming recognition and data relationship mining as two independent sub-tasks. However, this pipelining approach has the following disadvantages: first, errors in the data naming recognition module will spread to the stage of data relationship mining, thus affecting the accuracy of data relationship mining; Secondly, the pairings and combinations of data in the data relationship mining stage produce a lot of redundant information. Therefore, how to jointly extract data and its semantic relationship will become the focus of future research.

3.5. Data fusion and exchange technology

Data fusion exchange technology is mainly used to eliminate the situation that entities extracted from different data resources have the same meaning in the process of data name recognition and data relationship mining. Firstly, the heterogeneous data entities are integrated with resources. Then the multi-source relationship is integrated based on data mining and other technologies. Finally, the heterogeneous data resources are constructed into a unified data resource graph to ensure the quality of the data resource graph construction.

(1) Heterogeneous data integration

The business system can be optimized and adjusted by using the technology of data center and entity relationship extraction, so as to make the relationship between business and business exist. Finally, the data resources between different business systems can be shared and automatically synchronized in real time under the condition of no intervention.

(2) Entity disambiguation

It is difficult for entity recognition results to be directly added to the data resource graph due to the phenomenon of polysemy or multi-word polysemy in natural language. Therefore, for data entities in different business areas of energy and power systems, data mining algorithms such as clustering, entity linking and collaboration are used to disambiguate the results of entity recognition.

(3) Entity alignment

Entity alignment is to judge whether two or more entities in the same or different data sets point to the same entity, then gather entity names with the same reference together and realize the alignment and resource fusion of multiple entities. Firstly, extract the open link data and the set of data entities with the same name in the energy and power industry. Then, algorithms such as entity alignment method are propagated based on similarity, and the upper concepts of these data entities are compared. If they have the same upper concepts, they are combined into one entity [16].

4. Analysis on the application prospect of data resource graphs in the context of energy Internet

In the context of energy Internet, the application of grid data resource graph is mainly reflected in the semantic fusion of multi-source heterogeneous data, intelligent processing of big data and the construction of enterprise data center, as shown in the figure below.
4.1. Power grid data resources semantic integration

With the construction of the energy Internet, the data resources collected by the platform layer are faced with such problems as decentralization, fragmentation, sharing difficulties and lack of connection, etc., and the data acquisition mode presents the characteristics of cross-end, cross-source and cross-mode. Therefore, the semantic aggregation of data resources has gradually become a research hotspot in the field of power system.

The data resource graph has a strong inclusion ability, and can obtain power grid dispatching data from various aspects. The related data mainly come from the basic database (namely relational database) provided by relevant departments of power grid, and the data collected by operation scheduling and text images (non-relational data). Relational databases can be graphed directly to RDF, further converted into RDF files and imported into RDF databases, which then constitute the data entities in the data resource graph and are uploaded to the network layer. For non-relational database, on the basis of the semantic search system in the domain of electric power operation demand, combined with technical personnel operation target, first to extract relevant information in the database, the data into an XML file, then to participle of unstructured information, make it more specific and easy to understand, the last on the basis of relevant properties, relationships and entities to convert the file format for RDF file. The child nodes in the XML file can generate entities and object attributes as traversal child nodes, presenting a recursive state in turn. With the help of JenaAPL technology, relevant knowledge acquisition targets are implemented, and the obtained files are stored in RDF database, that is, semantic search data storage system in power field, laying a foundation for the construction of data resource graph.

4.2. Power grid big data intelligent processing

Power grid big data is the foundation of energy Internet construction. Through the data resource graph, the massive and complicated big data in the field of energy and power systems can be integrated into a semantic data network, to solve the problem of data convergence and fusion, and to optimize the process of data value acquisition.

From the perspective of data convergence and fusion, the problems existing in the convergence and fusion process of the current power grid big data are as follows: the lack of unified standards and norms, the difficulty in the dynamic change of data mode, the difficulty in semantic fusion of multi-source
heterogeneous data, "data island" and "data chimney" and other problems are still serious. Therefore, there is an urgent need for a flexible, scalable and intelligent adaptive data model to deeply integrate existing data into multiple dimensions. As a lightweight data pattern, data resource graph has the ability of semantic association and dynamic extensibility, which can realize the unified modeling and management of multi-source heterogeneous data to a certain extent.

From the data analysis and mining level, because of the lack of context and context domain knowledge, the energy power system in the field of text, images, video and other unstructured data processing, the traditional text mining algorithm (text categorization algorithm, latent semantic analysis algorithm, etc.) as well as the deep learning models such as CNN, RNN analysis effect is poor. Therefore, energy and power system domain knowledge provided by grid data resource graph is needed to further improve performance.

In addition, the power grid big data analysis and mining technology under the guidance of the current "data-driven" concept can well deal with the problems with "correlation" characteristics. However, it can't solve the problems of "causal relationship" such as causal judgment, retroactive reasoning and decision planning. The reason is that the "data-driven" approach relies heavily on the characteristics of the data itself and lacks the integration of domain knowledge and expert experience.

Therefore, through the construction of grid data resource graph, firstly, it can provide prior knowledge to analyze and mine power grid big data; secondly, it can enhance the learning ability of machine learning algorithm and improve its cognitive level, and finally realize the intelligent processing of grid big data.

4.3. Facilitate the construction of enterprise-level data center of power grid companies

At present, the systems of the state grid corporation of China are independent of each other, so there is a lack of data sorting and connection, a unified database and the correlation between data and intelligent recommendation function. Therefore, it is of great significance for power grid enterprises to improve the level of data service to build a data center in the energy Internet platform layer.

Based on the unified data center of the whole business and according to the requirements of data sharing and analysis application, the data center accumulates and improves the common data service capability to meet the requirements of data sharing, analysis mining and integration across different professional levels and vertical levels. The full-service unified data center is a platform for modeling, collecting, unified storage management, and analysis applications of structured, measured, and unstructured data for the power grid. However, there is no correlation between the data, and it is impossible to cross the disciplines to collect and store grid data. At the same time, because the data is not well expressed mathematically, data resources cannot be intelligently analyzed and managed.

Therefore, the data resource graph is applied to the entire business unified data center and the association between data is established, so as to establish a semantic-level interoperable unified data service. Finally, the efficient and intelligent query of grid data and the cross-business connectivity of enterprise data will be achieved, which will help the construction of a platform for enterprise data, thereby improving the company's economic efficiency. Internally, it can support business distribution, new-generation marketing services, multi-dimensional lean management, digital audit, smart supply chain, online power grid, comprehensive digital management of the whole process of infrastructure construction, power supply service command and other power grid businesses while externally, it can support for comprehensive energy services, Internet of vehicles, e-commerce, supply chain finance, virtual power plants and other emerging businesses.

5. Summary and outlook

Deep learning and knowledge graph are the latest achievements of the research paradigm of artificial intelligence symbolism and connectionism. The two technologies are deeply integrated with each link of the energy and power system, and the grid data resource graph is constructed to provide technical support for the company to establish the enterprise-level data center. Thus, it can solve the management
and fusion exchange of multi-component and heterogeneous data standardization under the massive data faced by power grid companies.

Although data resource graph, as the core foundation of integrating multiple and heterogeneous data resources of power grid, has a broad application prospect in the construction of energy Internet, it is still in its infancy in basic theoretical research, platform development and application demonstration. Therefore, the future development of data resource graphs should be based on multi-dimensional perspectives such as data modeling, resource management, data navigation, learning cognition, and databases. Deeply study of the basic theories and methods of multi-heterogeneous data resource graphs of power grid companies, and then construct a data resource graph including data ontology modeling, data naming and identification, data relationship mining, and data fusion exchange. Further play a role in actively promoting the demonstration application of data resource graphs in Taichung in company data, so that the implementation and implementation of data resource graphs in the field of energy and power systems is effectively promoted.

Acknowledgments
This paper was supported by the Science and Technology Project of SGCC (Research on key technology of data center design and realization (SGFJXT00WLJS1900278)).

References
[1] SHEN Xin, CAO Min, ZHANG Jingjiang, etc.. Research and Application of Panoramic Information Architecture Model of Power Internet of Things Devices Based on Multi-source Heterogeneous Data Aggregation Technologies [J]. Southern Power System Technology, 2016,10(2):48-55.
[2] JIN Zhengshu, WANG Yanbin, FAN Xiaoling. Share of heterogeneous database based on XML in administration of electric power enterprise [J]. Journal of Northeast China Institute of Electric Power Engineering(Natural Science Edition), 2005,25(1):76-79.
[3] WANG Tong, YE Youming. The Distributed File System of a Mass Heterogeneous Data Storage for The Power Industry [J]. Electrical Engineering, 2015(1):95-97,101.
[4] FENG Ying, TAN Yonggui, HE Anhong. Comprehensive Monitoring and Analysis Platform for Power Grid Information Based on Real-time Database [J]. Electrotechnical Application, 2015,34(Supplement 2):623-628.
[5] LI Juanzi, HOU Lei. Summary of Knowledge Graph Research [J]. Journal of Shanxi University(Natural Science Edition), 2017(3):454-459.
[6] Liu Qiao, Li Yang, Duan Hong, etc.. Knowledge Graph Construction Techniques [J]. Journal of Computer Research and Development, 2016(3):582-600.
[7] Xv Zenglin, Sheng Yongpan, He Lirong, Wang Yafang. Knowledge Graph Technology Overview [J]. Journal of University of Electronic Science and Technology of China, 2016(4):589-606.
[8] Zhao Jun. Knowledge Graph [M]. Beijing:Higher Education Press, 2018:1-20.
[9] LIU Bochong. Introduction to "Theory and Application of Chinese Domain Ontology Automatic Construction" [M]. Journal of Ningbo University(Liberal Arts Edition), 2015:27-35,44-45.
[10] WEI Shunping, HE Kekang. Semi-automatic Building Approach of Domain Ontology Based on Text Mining: A Case Study of Building Instructional Design Domain Ontology [J]. Open Education Research, 2008(5):95-101.
[11] Morwal S, Jahan N, Chopra D. Named Entity Recognition Using Hidden Markov Model (HMM) [J]. International Journal of Natural Language Computer, 2012 (4): 15-23.
[12] Ahmed I, Sathyraj R. Named Entity Recognition by Using Maximum Entropy[J]. International Journal of Database Theory & Application, 2015 (8): 43-50.
[13] Huang Z, Xu W, Yu K. Bidirectional LSTM-CRF Models for Sequence Tagging [EB/OL]. [2019 -06 -12]. https: //arxiv.org/pdf/1508. 01991v1.pdf.
[14] Ma X, Hovy E. End-to-end Sequence Labeling via Bi-directional LSTM-CNNs-CRF [EB/OL]. [2019-06 -12].https: //arxiv.org/pdf/ 1603.01354.pdf.
[15] WANG Dongbo, GAO Ruiqing, SU Xinning, etc. Automatic Extraction and Analysis of Data Scientific and Technical Literacy for Informatics Curriculum [J]. Information Studies: Theory & Application, 2018(12): 61-67.

[16] LACOSTE – JULIEN S, PALLA K, DAVIES A, et al. SIGMa: simple greedy matching for aligning large knowledge bases [C]// Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2013: 572 – 580.