Design and Implementation of Digital Platform Based on Standard Knowledge Map in Oil and Gas Exploration and Development Field

Shan Huang*, Xuesong Su[a], Wangkai Yue[a], Yangxin Xin[b], Wangyi Fei[c]

Technology Inspection Center of Shengli Oilfield, SINOPEC, Dongying, Shandong,China

* Corresponding Author Email: huangshan.slyt@sinopec.com, [a]wangkaiyue.slyt@sinopec.com, [b]yangxinxin.slyt@sinopec.com, [c]wangyifei524.slyt@sinopec.com

Abstract. This paper introduces the current situation and problems of standardization development in the field of oil and gas exploration and development in China. To solve the problems of large amount of standard knowledge, large professional span and complex level, the concept connotation of ontology and knowledge graph is clarified. A triplet core data model of "standardized object (product) -style (paragraph structure) -index (attribute)" is proposed for oil and gas exploration and development. Based on this, the construction idea, main function, and future direction of digital platform of standard knowledge map in oil and gas exploration and development are studied.

Keywords: oilfield field, knowledge graph, service platform, design and implementation.

1. Introduction

The development of the oil and natural gas industry has formed a comprehensive industrial system covering over a dozen specialties, including petroleum exploration, development, oil and gas gathering and transportation, and storage and transportation. With the rapid growth of China's petroleum and natural gas industry, the scale of enterprises continues to expand, and the difficulty of oil and gas exploration and development is increasing. The safety situation in the field of oil and gas exploration and development is quite serious, facing unprecedented challenges and competition. At the same time, more and more safety, health, and environmental issues are being exposed. The safety and environmental protection issues in the field of petroleum and natural gas exploration and development, including the objects faced, technical requirements, new technologies, and the application of new technologies, are becoming increasingly prominent. Currently, public awareness of safety and environmental protection is constantly increasing worldwide, and the attention to safety accidents and environmental issues has reached an unprecedented level. The American Petroleum Institute (API), which is a standard-setting organization in the US oil and gas industry, has a safety, health, and environmental department that is responsible for developing and revising health, safety, and environmental standards. The exploration and production department and refining department also separately develop safety and environmental protection standards in their respective areas. China's work in standardizing safety production technology has formed several major systems of safety and environmental protection standards, including design standards, safety production equipment/tools standards, production process safety and hygiene standards, protective equipment standards, management standards, and safety technical regulations.

Studies have shown that accidents occur frequently when not following standard operating procedures, with human error accounting for about 70% of the total number of accidents. The petroleum and petrochemical industry are a high-risk industry [1]. Once a large or major accident occurs, it will bring serious disasters to society, cause huge financial losses to enterprises, and have a negative impact on the corporate image.

Safety and environmental protection issues in oil fields are two important issues in oil field production that are also of great concern to society. The relevant departments of the oil field are enhancing their understanding of these two issues, using the correct strategies and effective methods to make digital standards play a decisive and regulatory role in oil field environmental protection and safety work, which is the basic starting point of this study.

2. Standardized data status and issue analysis in the field of oil and gas exploration and development

2.1. The current number of existing standards and regulations in oil fields is large, but the level of digitization is not high

Currently, the standard system used in the field of oil and gas exploration and development in China includes platforms such as the dynamic management system of standards, the system for standard formulation and revision, the oil field standard query system, and the technical supervision management platform [2]. The historical standard data of each platform is stored independently, and the standard specifications and formats are not unified, which brings inconvenience to the uniform use and centralized management of standards. In addition, the existing standards of enterprises, national, industry, and local standards still exist in the form of non-structured electronic documents. The digitalization of core elements of standards, such as standard content, clauses, and indicators, is not high, and there are significant shortcomings in guiding oil field environmental protection and safety production activities.

2.2. The traditional methods of obtaining standards are outdated, and subjective activities by individuals lead to significant uncertainty and risk in standard usage

Oil fields have numerous and widespread safety risk points, and the management of environmental protection and safety in the field of oil and gas exploration and development is a typical knowledge-intensive work. However, this knowledge is stored in various types of documents, such as standard specifications, construction organization designs, safety technical solutions, accident investigation reports, and various other technical and management materials. In the process of environmental protection and safety management in oil fields, a large amount of specialized knowledge is needed. However, due to the large volume, variety, and wide sources of the data, it is difficult to extract valuable knowledge quickly and accurately from these materials.

2.3. Inadequate guidance from standards for safety inspection work

Currently, on-site operational implementation and supervisory personnel in oil fields often need to carry many paper-based standard documents and regulatory files with them during their daily work. After discovering problems, a considerable amount of effort is required to search for relevant information. However, existing search engines still mainly rely on file names or numbers for retrieval. It requires a significant amount of manpower and time to interpret the knowledge contained in paper or electronic standard documents, regulatory files, and laws and regulations. There are issues such as incomplete resources for retrieving standards and regulations, outdated search methods, low retrieval rates, and low retrieval accuracy, which significantly affect the widespread application and promotion of standards, regulations, and relevant laws and regulations.

Considering these issues, this paper stores national, industry, local, and group standard data in the field of oil and gas exploration and development in databases and file storage, and builds a standard catalog database in the data management system through the integration of data from multiple heterogeneous sources. Subsequently, the standard data is identified for content, and a knowledge graph of standard content is constructed. The standard knowledge graph contains standardized objects, indicators, limitations, and formats. Ultimately, digital research services for knowledge graphs in the field of oil and gas exploration and development, such as catalog retrieval, full-text search, and indicator display, are provided on the PC end.

3. Building a digital platform for standard knowledge graph in the field of oil and gas exploration and development

3.1. Ontology

The concept of ontology originated in the field of philosophy, used to describe the essence of things. After the term ontology was coined, scholars have consistently researched and discussed its concepts and implications. In 1998, scholars such as Student provided a clearer and more comprehensive explanation of the characteristics of ontology:

"Knowledge ontology is a clear, formal, and shareable normative description of conceptual systems," and further pointed out that a knowledge ontology is "the abstraction and description of domain knowledge norms, and the method to express, share, and reuse knowledge."

If each knowledge domain is abstracted into a set of conceptual systems, further specified as a vocabulary to represent, including the clear definitions of each word, relationships between words (e.g., representation, classification, part-whole, and reference relationships), as well as statements of some axiomatic knowledge in that domain, and if a consensus can be reached among experts in that domain to share this vocabulary, all of the above constitute a "knowledge ontology" for that knowledge domain.

Finally, for computers to understand and process, a specific coding language (e.g., RDF/OWL) is needed to explicitly represent the system (vocabulary, vocabulary relationships, relationship constraints, axioms, inference rules, etc.). In this sense, knowledge ontology has become a tool for extracting, understanding, and processing domain knowledge and can be applied to any specific scientific and professional field. In short, ontology is the foundation for building a knowledge graph.

3.2. Knowledge graph

A knowledge graph is an abstract representation of elements in a specific domain of the real world and their interrelationships, a structured knowledge representation model that is computable and understandable by computers. Knowledge graphs abstract "elements" in a domain as "entities" in an object-oriented manner, representing them as nodes in the graph, and abstract the inherent connections between two elements as "binary relationships," represented as edges (connecting two nodes) in the graph. The names of entities and relationships are labeled to form triple structures of two entities and a relationship (as shown in Figure 1).

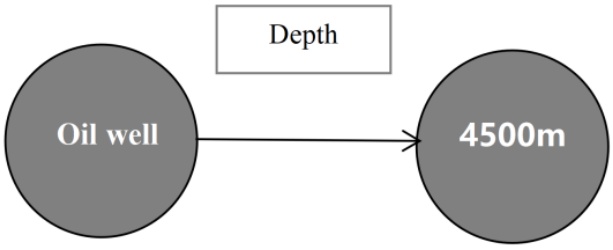

Figure 1. Triple structures of knowledge graph

Entities refer to the people (e.g. oil field workers), things (e.g. requirements for controlling atmospheric pollution emissions), objects (e.g. oil wells), attribute values (e.g. 4,500 meters), or concepts (e.g. the 14th Five-Year Plan for oil fields) that exist in the objective world. Relationships refer to the connections between entities (e.g. cause-and-effect relationships) or attributes (e.g. depth).

A triple consists of two entities and the relationship between them [3]. For example: Entity 1 is an oil well, the attribute is depth, and Entity 2 (which is an attribute value) is 4,500 meters.

3.3. Creation approach

As for the creation approach, in response to challenges such as fragmentation of standard documents in the oil and gas exploration and development field, difficulty in associating data between standards, and complex data structures making it hard for machines to understand, the approach combines the

characteristics of the petroleum industry's exploration and development field with the attributes of environmental protection and safety. This involves using technical materials such as environmental safety standards and regulations as the underlying data, designing a methodology for organizing knowledge about oil field environmental protection and safety standards, and clarifying the display format of the knowledge. It also includes researching semi-automated standard knowledge organization techniques based on natural language processing and machine learning, as well as establishing the interrelationships between the elements of the knowledge system. This results in forming a knowledge graph of standard regulations for oil field environmental protection and safety and using this to build a digital platform based on a knowledge graph of standards in the oil and gas exploration and development field. This establishes a solid technical foundation for extending the application of this platform to other fields. The creation approach for the digital platform based on the knowledge graph of standards in the oil and gas exploration and development field mainly involves the following four aspects.

3.3.1 Development of organizational methodology for oil field environmental protection and safety standards knowledge based on knowledge ontology theory

The organizational methodology for standard knowledge is the basis and reference for identifying entities and extracting relationships among standard content in the knowledge graph: 1) Combining the oil field environmental protection and safety management system and the oil field standard system, design and create a standard knowledge classification system suitable for the oil field environmental protection and safety field, covering various organizational dimensions such as work scenarios, business processes, and application equipment. 2) Based on this, combined with the business dimensions of the oil field (such as exploration and development, surface engineering, public engineering, and offshore engineering), and based on knowledge ontology theory, adopt methods such as thesauri to research the core data model of triples applicable to the oil and gas exploration and development field, namely, standardized object (product) -style (paragraph structure) -index (attribute), such as: "Light petroleum products—Acidity—Neutralizing milligrams of potassium hydroxide required for 100ml of light petroleum products" (GB/T 258-2016 "Determination of Acidity in Light Petroleum Products"). In this model, synonym and hypernym relationships need to be established for both products and specimens, and the indicators include indicator items, indicator values, units of measurement, and limited categories. This achieves fragmented documents and forms an organizational methodology for oil field environmental protection and safety standards knowledge based on knowledge ontology theory and using methods such as thesauri.

3.3.2 Development of semi-automatic standard knowledge organization technology based on natural language processing and machine learning

Different from general knowledge graphs, the knowledge graph applicable to the environmental protection and safety field of the oil and gas exploration and development domain is specific to this field, and has strong depth, completeness, and relevance, with numerous entity attributes and industry significance. Typically, the construction of the data model for industry domain knowledge graphs is undertaken by experts in the field. However, the experts have limited understanding of knowledge graphs in computing, leading to long construction periods, high labor costs, and difficulties in initiating knowledge graph construction. To significantly reduce the construction cost of industry knowledge graphs during initiation, semi-automatic processing of semi-structured and unstructured data is carried out: 1) Using manually processed structured data as a training set, applying a machine learning framework to achieve automatic entity and relationship labeling for semi-structured data, and constructing a standard knowledge graph. 2) Using manually constructed vocabularies and syntax rule paradigms as a basis, achieving entity identification and disambiguation, relationship labeling for unstructured data, and constructing a standard knowledge graph. Subsequently, experts summarize key information, optimize labeling results through iteration, and provide the labeled results for expert selection and judgment, thereby accelerating the knowledge graph construction process.

3.3.3 Establishing a standard knowledge base framework based on relationship networks and knowledge graphs

After defining and clarifying the research objectives, types of objects, and specifying object templates, we study knowledge extraction techniques based on domain pre-training models to develop a knowledge architecture primarily based on tree structures and auxiliary attribute classification, as well as a knowledge association network primarily based on triplets [4]. By researching and applying techniques such as topic modeling, label relation mining, semantic representation, and multi-level knowledge integration, we construct an integrated system for identifying and annotating knowledge objects at different granular levels, including phrase-level, sentence-level, and chapter-level, to meet the requirements for annotating and extracting general fine-grained knowledge objects and domain-specific knowledge.

3.3.4 Creating a digital platform for knowledge graphs in the field of oil and gas exploration and development

Addressing the practical scenarios of standard literature management, research, and utilization in the field of oil and gas exploration and development in the petroleum industry, we are creating databases of standard knowledge graphs at different levels. Ultimately, on the PC end, the platform will offer knowledge services such as standard knowledge retrieval, classification navigation, knowledge graph, and information recommendation. This will result in a digital platform for knowledge graphs in the field of oil and gas exploration and development, effectively addressing issues related to precision in standard literature search, intelligence in knowledge discovery, and accuracy in literature association recommendation.

4. Digitized platform for standard knowledge graphs in the field of oil and gas exploration and development

Knowledge graphs can be applied and developed in the process of digital transformation of standards to help users achieve more intelligent and efficient standard information processing and management. Specific functions include the following:

4.1. Search engine optimization

By building a digitized platform for standard knowledge graphs in the field of oil and gas exploration and development, search engines can better understand user search results and provide more accurate and useful information. The knowledge graph contains rich structured standard data and entity relationships, allowing the search engine to better understand user query intent and provide more accurate search results.

4.2. Content recommendation

Based on user behavior and preferences, using the digitized platform for standard knowledge graphs in the field of oil and gas exploration and development to recommend related content, improving user experience and satisfaction. The rich semantic information can expand search results, providing a better search experience.

4.3. Intelligent customer service

The digitized platform for standard knowledge graphs in the field of oil and gas exploration and development combines ontology, knowledge base, and knowledge graph, allowing intelligent customer service systems to answer user questions more quickly and accurately, improving the efficiency of standard information and intelligence service.

4.4. Data analysis

The digitized platform for standard knowledge graphs in the field of oil and gas exploration and development utilizes the relationship network of the knowledge graph for data mining and analysis, discovering potential patterns and trends to support decision-making.

4.5. Machine learning

Knowledge graphs can serve as input data for machine learning models, used to train models and improve prediction accuracy.

5. Outlook of future technological

Currently, the development of information systems based on knowledge graphs is slow. With the rapid emergence of technologies such as deep learning and neural networks, the digitized platform for standard knowledge graphs in the field of oil and gas exploration and development is also being propelled forward, and there is a vast research space. The future development direction of the digitized platform for standard knowledge graphs in the field of oil and gas exploration and development is as follows:

1)Based on neural network deep learning, the quality of the knowledge graph determines the experience of the standard digitized platform. Therefore, the problem of link prediction in the knowledge graph of the oil and gas exploration and development field cannot be ignored. For example, if there are missing entity relationships in the oilfield knowledge graph, it may lead to the inability to complete the closed loop in the oil and gas production process, and the quality of recommendations may decline due to the lack of factual relationships. Therefore, deploying knowledge graph reasoning and recommendation tasks in the same model is an effective way to enhance the quality of recommendations. In addition, a multi-task information system for oil and gas exploration and development is intelligent and promising, as it can release and alleviate the pressure of scarce production resources, making production tasks more convenient and efficient.

2)An intelligent oilfield management system based on knowledge graphs can achieve comprehensive monitoring, early warning, and operational optimization through modeling and analysis of various data, equipment, and processes. This will significantly improve the production efficiency and safety of the oilfield. Establishing a knowledge graph in the field of oil and gas exploration and development can aggregate and integrate various data, literature, and expert experience, thereby supporting more precise oilfield decision-making and risk assessment using artificial intelligence algorithms. As the energy Internet of Things continues to be developed, knowledge graphs can play a crucial role in connecting various aspects, such as optimizing efficiency and intelligent scheduling in multiple stages, from natural gas transmission to processing and terminal applications. The establishment of an industry-level oilfield knowledge graph will also promote the standardization and regulation of the oil and gas exploration and development field, further enhancing industry efficiency and quality levels.

References

[1] Liuwu Li, Yuqi Bu, Yi Cai*. Bottom-Up and Bidirectional Alignment for Referring Expression Comprehension. The 29th ACM International Conference on Multimedia (ACM MM 2021).

[2] Changmeng Zheng, Yi Cai*, Guanjie Zhang, Qing Li: Controllable Abstractive Sentence Summarization with Guiding Entities. COLING 2020: 5668-5678.

[3] 29.Xin Wu, Yi Cai*, Kai Yang, Tao Wang, Qing Li: Task-oriented Domain-specific Meta-Embedding for Text Classification. EMNLP 2020: 3508-3513.

[4] Junsheng Kong, Zhicheng Zhong, Yi Cai*, Xin Wu, Da Ren: TSDG: Content-aware Neural Response Generation with Two-stage Decoding Process. EMNLP (Findings) 2020: 2121-2126.