Research on the Construction Method of Knowledge Ontology Facing the Field of Substation Maintenance

Bowen Zhang1*, Ping Qian2, Cheng Li2, Fei Gao1, and Ning Yang1

1China Electric Power Research Institute, Beijing, China
2Electric Power Research Institute of State Grid Zhejiang Power Company, Zhejiang, China
*Corresponding author’s e-mail: zhangbowen@epri.sgcc.com.cn

Abstract. Starting from the definition and classification of knowledge, this paper gives the definition and classification of substation maintenance knowledge. Based on the "seven-step method", it proposes a method of constructing knowledge ontology for narrow fields, including 9 steps, and establishes substation maintenance knowledge Ontology, including 13 concepts and 60 attributes, lists the concepts and attribute descriptions of the ontology, laying the foundation for the next step of building a knowledge navigation map.

1. Introduction
Ontology originated in philosophy, which refers to the interpretation and explanation of objective existence system. After a thorough study of ontology in 1993, Gruber and others gave a clear, comprehensive and general concept, which they thought was "a clear formal specification of shared conceptual model ". In recent years, ontology has been widely used in related fields and has developed rapidly. Under this background, ontology construction has become a hot topic. The ontology construction method has not yet a complete system, but for the ontology construction principle, Gruber proposed in 1995 five rules are relatively influential. Domain ontology is a kind of ontology by defining the elements such as class, instance, attribute, relation, axiom and so on, which depicts the class and instance in a certain field and their hierarchical relationship, and induces and abstracts the domain knowledge [1-2].

Starting from the definition and classification of knowledge, this article gives the definition and classification of substation maintenance knowledge. It starts with the research ideas and methodology of the general subject theoretical system, and based on "logical starting point theory" and "problem system theory". The problems to be solved in the substation maintenance, study the composition and logical structure of the substation maintenance knowledge system, and then give the substation maintenance knowledge system framework.

2. Knowledge Ontology Construction Method
At present, almost every ontology system development will form the corresponding ontology construction method. Because the subject areas involved are different, the specific objectives of ontology realization are different, and the content of comparative research involved in the construction of ontology is not the same. Therefore, there is no set of system and complete theory for ontology construction method. For the foreign domain ontology construction methods for systematic analysis has been more mature [3-4].
2. 1. IDEF5 law
IDEF5 is an ontology description acquisition method, which provides a structured method available for the development of high-precision domain ontology. The defect of this method is that there is no intention of cyclic development. IDEF5 ontology development process includes the following five activities: 1 Organize and carry out activities. 2 Data acquisition. 3 Data analysis. 4 Initial ontology development. 5 Perfection and verification of ontology.

2. 2. Skeleton method
Skeleton method is a summary of experience in Uschold and King(1995) development EO (Enterprise Ontology, University of Edinburgh Enterprise Modeling Ontology). This method clearly describes the basic process and guidelines of ontology development, which has important guiding significance for the current ontology development practice, and the existing problem is that the evolution of ontology is not clearly proposed. The specific content includes the following four aspects: First, determining the reference purpose and scope of the ontology. Second, building ontology. Third, evaluation of ontology. Fourth, avoiding barriers to knowledge sharing. The flow is shown in Figure 1.

![Figure 1. Skeleton Method Flow Chart](image)

2. 3. TOVE law
TOVE method, also known as evaluation method, is a summary of experience in developing TOVE engineering ontology such as Gruninger and Fox. This method does not directly construct the knowledge logic model described in ontology form, but first establishes the informal description of ontology and then formalizes the description. TOVE steps include the following aspects: First, get the incentive plot. Second, the ability to articulate non-formalization. Third, standardization of terms. Fourth, identification of formal capacity issues. Fifth, Determination the axioms and definitions in formal language. Sixth, evaluation the integrity of the ontology. Its flow chart is shown in Figure 2.

![Figure 2. TOVE Flowchart](image)

3. Method of Building Knowledge Ontology Based on "Seven-step Method"
In order to make up for the deficiency of "seven-step method" in knowledge ontology evaluation and concept attribute collection, this paper proposes a knowledge ontology construction method based on "seven-step method", which includes 9 steps, as shown in figure 4. Of these, steps 1, 4, 6-8 are identical to the seven-step approach, and steps 2, 3, 5 and 9 are described in detail below. 2 4
Determine the professional field and category of the ontology
Collect relevant information about domain knowledge
Determine the principles of ontology construction

Define classes and class hierarchy
Extract conceptual attributes from domain knowledge data
Reuse existing ontology

Define the attributes of the class
Facets that define attributes
Evaluation

Figure 3. Knowledge ontology construction process based on "seven-step method"

Step 1 is to determine the professional field and scope of the ontology. We must first clarify a few basic questions:

(1) Which professional field will the constructed ontology cover;
(2) The purpose of applying the ontology—To better dig deep information in the field;
(3) What types of questions can the information in the ontology answer;
(4) Who are the users and system maintainers of the ontology? The answers to these questions can be adjusted at any time with the deepening of the ontology design process, but in any specific period, they are helpful to limit the scope of the model, so they need to be relatively stable.

Step 2 Collect domain knowledge-related information: In view of the narrow scope of the field of substation overhaul, in order to be able to comprehensively list the domain-related concepts and attributes, you can collect all the information related to domain knowledge, including structured data dictionary, unstructured (semi-structured) text data, etc. In order to collect domain knowledge comprehensively, it is necessary to establish domain knowledge system first and then collect core information according to knowledge system.

Step 3 determines the principle of ontology construction: ontology application scene is closely related to ontology structure, which directly determines the hierarchy depth of ontology, granularity of concept, whether a term is attribute or concept, etc. The above three problems should be defined according to the ontology application scenario.

The hierarchy depth of the ontology is several layers.
The granularity of concepts and attributes.

Determine whether the term is a concept or an attribute based on the complexity of the term in the application scenario.

Step 4 is to examines the possibility of reusing existing ontology. If your system needs to interoperate with other application platforms, and this application platform is combined with a specific ontology or controlled vocabulary, then reuse of existing ontology is the most effective method. Many ontology have electronic versions and can be input into personal ontology development systems. Even if a knowledge expression system cannot directly work in a special format, it is not difficult to convert the ontology of knowledge from one format to another. You can find many ready-made ontology libraries on the Web.

Step 5 is to define the class and hierarchy. There are several feasible ways to perfect a hierarchical system: Top-down approach: Start with the biggest concepts in a certain field, and then refine these concepts. Bottom-up approach: Start with the definition of the smallest classes at the bottom, which are the minutaie of this hierarchy, and then organize these refined classes under a more comprehensive concept. Comprehensive method: combine the above two methods. First define a large number of important concepts, and then appropriately summarize and deduce them. Then associate them with some intermediate concepts.
Step 6 Extract Conceptual Attributes from Domain Knowledge Data: Part of speech tagging of data dictionary collected in step 2, word segmentation and part of speech tagging of text data.

Step 7: Define the facets of attributes (Facets) An attribute may consist of multiple different facets. Facets are used to describe the type of value (Value Type), allowable values (Allowed Values), the number of values (cardinality set potential, cardinality) and other characteristics related to attribute values.

Step 9 evaluation: the ontology inconsistency is judged by collision tree debugging algorithm based on Reiter diagnosis theory.

4. Construction of substation maintenance knowledge ontology

According to the knowledge ontology construction method based on "seven-step method" proposed in Chapter III, this chapter establishes the knowledge ontology of substation maintenance.

Step 1: Identify areas and areas of expertise in knowledge ontology

First, many reference documents, including unstructured text, are needed to make the maintenance plan automatically. Secondly, the automatic and intelligent work of the maintenance plan is carried out to improve the efficiency of the maintenance decision while reducing the oversight of the maintenance demand caused by subjective factors.

Step 2: Collect relevant information on domain knowledge

The knowledge system of substation maintenance is shown in Chapter 3, and the relevant materials are mainly stored in books, standards, systems and information systems, and the specific collection data is shown in Table 2.

Step 3 Determination of Ontology Construction Principle

The application scene of substation overhaul body includes intelligent retrieval of substation overhaul knowledge, intelligent question and answer and intelligent formulation of overhaul plan. The following principles for ontology construction are further clarified.

1. The hierarchy depth of the ontology is several layers. For intelligent retrieval, intelligent Q & A application scenario, due to the more information involved, and considering the time complexity of retrieval algorithm and the user's trust in the information provided, it is suggested that the depth of ontology level should not exceed three layers. It is suggested that the depth of ontology level should not exceed five layers because the requirement of the scene is specific and clear.

2. The granularity of concepts and attributes. The granularity of concepts and attributes can be divided into three categories: coarse, medium and fine. Among them, the coarse particle size is paragraph level, the grain size is sentence level, and the fine particle size is phrase lexical level. For intelligent retrieval, intelligent Q & A application scenarios, according to user needs coarse granularity can be. For the maintenance plan intelligent development scenario, due to the need for further recommendation algorithm analysis, it is recommended for fine particle size.

3. Determine whether the term is a concept or an attribute based on the complexity of the term in the application scenario.

Terms that are generally unnecessary for further subdivision or association with other concepts can be used as attributes, including numerical parameters, non-important terms, etc.

Step 4: Examine the possibility of reusing existing knowledge ontologies

At present, the knowledge ontology in power field is still in the research stage, and there is no ontology design which is recognized as mature or has good application effect.

Step 5: List important terms for ontology

Some terms in the field of substation maintenance have been given in the glossary of power subject words and related standards. In order to ensure the integrity of the terms covered by the ontology, the remaining data need to be processed, and the terms in the field of substation maintenance are refined and completed. Because the data dictionary in the information system has already standardized and stored the terms (fields) in the relevant fields, the terminology of structured data can be sorted out. The following is mainly aimed at the field of substation maintenance unstructured data terminology extraction methods are introduced.
(1) Text preprocessing

On the basis of the word list of discontinuation, Combined with the characteristics of the text, the invalid terms \n \t (, ), s, \\d, a-z, A-Z have been added. At the same time, a low-frequency word screening strategy was established for low-frequency words that appeared only once, Deactivating words and low frequency words were removed, The substation maintenance vocabulary with 1540 words, as shown in Table 1. Of which 773 are nouns, Including nouns, place names, nominal morphemes, names, verbs, other special names, nominal words, etc. Verbs 548, Including verbs, nominal verbs, verb elements, auxiliary verbs, etc. Adjective 77, Including adjectives, auxiliary words, nominal words, etc. The number of specific parts of speech is shown in Table 1.

Table 1. substation maintenance vocabulary

| No. | Words and phrases   | POS | Name of part | No. | Words and phrases   | POS | Name of part |
|-----|---------------------|-----|--------------|-----|---------------------|-----|--------------|
| 1   | Annex   n           | Nouns | 11 | Clean   z          | State word |
| 2   | Direction n        | Nouns | 12 | Inhalation n     | Nouns |
| 3   | Boxes   n           | Nouns | 13 | Article 5 n        | Nouns |
| 4   | Tolerance a        | Adjectives | 14 | Anti-freezing v    | Verb |
| 5   | Stopping v         | Verb | 15 | Due v             | Verb |
| 6   | Slip down v        | Verb | 16 | Pollution collection v | Verb |
| 7   | Filling pressure v | Verb | 17 | Dusty i           | Idioms |
| 8   | Object n           | Nouns | 18 | Drive tide n      | Nouns |
| 9   | Part II n          | Nouns | 19 | Thickness n       | Nouns |
| 10  | Benchmarking n     | Nouns | 20 | Vacuum n          | Nouns |
|     |                     |     |              |     |                     |     |              |

Table 2. substation maintenance vocabulary statistics

| No. | Name of part | Quantity |
|-----|--------------|----------|
| 1   | Nouns        | 644      |
| 2   | Verb         | 492      |
| 3   | Adjectives   | 69       |
| 4   | Adverbs      | 57       |
| 5   | Name verb    | 51       |
| 6   | Azimuth      | 39       |
| 7   | Names        | 33       |
| 8   | Quantifier   | 33       |
| 9   | Quantity     | 20       |
| 10  | Distinction  | 17       |
|     |              |          |

(2) Manual audit

Because there may be errors in the words extracted by automatic participle, it is necessary for professionals to further review, delete invalid words, modify wrong words, and supplement missing words.

Step 6 Defines the hierarchy of classes and classes

(1) Classification of lexical conceptualization

Since the words in Table 1 are mixed with the terms of entity level and concept level, it is necessary to further conceptualize and classify the above words.

(2) Manual audit

After the classification of lexical conceptualization is completed, professionals are required to merge or delete the concepts corresponding to less vocabulary, check whether the concepts review the professional application habits, etc. The audited concepts include 13 concepts of organization, power company, manufacturer, phenomenon, substation, equipment, components, parts, parts, switches, transformers, overhaul, personnel, measures, etc., as shown in figure 4.
Figure 4. Concept and relationship between maintenance and maintenance of substation

Step 7 Defines the properties of the class
According to the set of attributes after lexical conceptualization, the corresponding attributes are associated to each concept, and the English name and specific description of the attributes are specified.

Step 8 Defines the partition of the attribute
defines the cardinality constraints for each attribute.

Step 9 evaluation: ontology inconsistency judgment needs entity association, knowledge base is established, not carried out.

5. Conclusion
The substation maintenance knowledge ontology is the basis for constructing the substation maintenance knowledge map, and the ontology construction should first clarify the construction goal of the ontology, the application purpose of the ontology and the professional field of the knowledge ontology, and clarify the knowledge system, content and scene of the research field. Starting from the definition and classification of knowledge, this paper gives the definition and classification of substation maintenance knowledge. Based on the "seven-step method", it proposes a method of constructing knowledge ontology for narrow fields, including 9 steps, and establishes substation maintenance knowledge Ontology, including 13 concepts and 60 attributes, lists the concepts and attribute descriptions of the ontology, laying the foundation for the next step of building a knowledge navigation map.

Acknowledgement
We thank the State Grid Corporation of Science and Technology Project Research and Application of Knowledge Navigation and Decision Optimization Technology for Substation Maintenance Based on Artificial Intelligence (grant number 5200-201918255A-0-0-00).

References
[1] Yue L.X., Liu W.Y. (2016). Intelligence Theory and Practice, 2016, 039(008):119-125.
[2] Manning, C. D., Prabhakar R. (2009) An Introduction to Information Retrieval. UK Cambridge, Cambridge University Press.
[3] Manning, C. D. (2015) Computational Linguistics and Deep Learning. Computational Linguistics, Vol. 41, No. 4, pp. 1-7.
[4] Devlin J, Chang M. W, Lee K., et al. (2018) BERT:Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv:1810.04805v1.