Study on Ontological Theory and practice in Literature and Documentation
----Taking Standard Documentation as an Example

Li Jing[1] Li Guo peng[2] Tian Fang[1]

1. China National Institute of Standardization，Beijing, 100191 China, tianfang@cnis.ac.cn;

2. Institutes of Science and Development, Chinese Academy of Sciences, Beijing, 100190 China

Abstract:

Objective: The goal of this study is to practice the application of ontology in documentation.

Methods: Literature research method; Website research method; System development method

Results: Based-on the introduction of ontological theory and foreign & domestic research trends, ontology practices in standard documentation were described in this article. The next-step work items were also represented.

Conclusion: It was indicated that the ontology-based knowledge navigation, auto-subject cluster, content-discovered and nature language retrieval would be given full scope to their potential/practical functions in standard documentation in future.

Keywords: ontology；standard literature；Knowledge organization system（KOS）

1 Introduction: Basic Theory of Ontology

The concept of ontology comes from the field of philosophy. The earliest to introduce the concept of ontology was in the field of artificial intelligence (AI). The introduction of ontology in the field of AI aims to address issues related to knowledge representation and organization. The concept and methods of ontology entered the field of knowledge organization and Library Science in early 2000. The purpose of introducing ontology is to solve problems such as information retrieval and resource organization.

1.1 Concept of Ontology

Gruber from Stanford University [1993] first proposed the definition of ontology [1], and later Dr. Borst Pim [1997] made some revisions to Gruber's definition. After merging the two definitions, it is an explicit formal specification of a shared conceptualization. [2]

The most widely used ontology definition is proposed by Dr. Michael Ushold et al. [1996] of Boeing: "It is a vocabulary of terms and some specification of their meaning. [3]

Through summarizing the viewpoints of domestic and foreign scholars, the author believes that although

C. Bai et al. (eds.), *Proceedings of 2023 China Science and Technology Information Resource Management and Service Annual Conference (COINFO2023)*, Advances in Economics, Business and Management Research 293,
https://doi.org/10.2991/978-94-6463-498-3_19

there are many expressions for the definition of ontology, the necessary conditions for it should also be reflected in the definition. The author believes that ontology is a clear and hierarchical specification of certain themes. It is a widely recognized formal knowledge representation system that includes a vocabulary (or title list/terminology list), in which all logical declarations are used to describe the meanings and relationships between these terms, that is, how they are related to other terms.

1.2 The Function of Ontology

Due to the fact that ontology can become the base for interoperability between different systems, building ontology should mainly address the following five issues:

(1) Reaching a common understanding and understanding of information organization structure among users or software agents. [1, 4]

Assuming there are several web sites that contain medical information or provide medical e-commerce services. If these web sites share the same underlying ontology, then proxy servers can extract and integrate these information resources from different sites. Proxy software can utilize these integrated information resources to answer user questions or provide data to users.

(2) Reusable professional domain knowledge.

For example, many models in professional fields need to describe the concept of time. These descriptions include concepts such as time intervals, time points (moments), and related time measurements. If a research organization develops such a detailed universal ontology, other research organizations can easily reuse it into their respective professional fields. To build a large ontology, it is also possible to integrate several existing ontologies. It is also possible to reuse a universal ontology framework, such as UNSPSC ontology, to extend and fill UNSPSC ontology (Skeleton) and apply it to other fields. [5]

(3) By using the form of assertions (formulas), academic viewpoints and hypotheses within the professional field are made clearer. A clear and formalized explanation of professional domain knowledge is helpful for new users who must understand the meaning of terminology in the field.

(4) By detaching domain knowledge from the environment of operations research and knowledge management, users can follow the necessary specifications and execute programs to achieve retrieval and query related functions, without having to be IT experts themselves. [6]

(5) Analyzing the knowledge system structure of the professional field.

In attempts to reuse and extend existing ontologies, standardized formal semantic analysis of terms is extremely valuable. [7]

1.3 Research progress in ontology

1.3.1 Foreign Countries' Research

From the 1990s to the present, in the past two decades, developed countries in Europe and America have continuously supported the construction and development of ontology based knowledge organization systems. In the past decade, greater emphasis has been placed on the development and application of knowledge organization systems in network environments, such as the development of WordNet (Elementary Ontology) by Princeton University, Classification Web by the Library of Congress (USA), Dewey Web by OCLC, and Ontolingua and Protégé ontology development tools by Stanford University, The National Library of Medicine of the United States has developed MeSH and UMLS (Unified Medical Language System). These knowledge organization systems have become essential infrastructure and tools for information development and utilization in the network environment, and are gradually moving from Western language to multilingual information processing. Its strategic intention to seize the dominant position in information processing is obvious.

Gene ontology is a widely used ontology in the field of bioinformatics, used to create shared biological information resources, allowing the industry to use common vocabulary and semantics to describe gene products, providing consistent descriptions of gene products in different databases. It mainly includes three branches: biological processes, molecular functions, and cellular components. GO contains a vocabulary of 90261 terms. [8]

UMLS is an integrated medical language system researched and developed by the National Library of Medicine in the United States since 1986. It consists of four parts: a super thesaurus, a semantic network, an intelligence source graph, and an expert dictionary, providing mappings between multiple thesauri, terminology lists, and thesauri.

In recent years, major publishers around the world have been fully promoting digital publishing, actively intervening in networked end-user services, and seizing the traditional service areas of scientific and technological literature and information institutions. Almost all digital publishers are actively launching intelligent retrieval, disclosure, data mining and analysis services for content objects, such as Google Book Search, Elsevier Scopus service, and Thomson Reuters Web of Knowledge Local service. The trend of their business strategy shift from mainly focusing on publishing and selling literature resources to providing independent knowledge organization systems and knowledge services is very obvious; Its strategic intention to gradually rely on the monopoly of resources and the expansion and improvement of service capabilities to meet user needs, seize the leading position in information processing, and achieve control over China's literature and information service market is obvious.

1.3.2 Domestic research

From 2006 to 2008, research hotspots in the field of ontology technology in China include: ontology tools, ontology learning, ontology libraries, ontology extraction, ontology reasoning, ontology translation, and ontology language. Domestic scholars and institutions tend to focus more on algorithms, matching, mapping, modeling, reasoning, and extraction research. In all relevant research, semantic analysis and semantic indexing are key tasks. After 2008, projects solely focused on ontology theory were no longer included in the scope of fund project applications. [9]

In the 2006-2009 support plan project of the Ministry of Science and Technology, "Common Technology Research on Rural Knowledge Ontology Construction" (2006BAD10A050104), the development of a large-scale ontology development environment (LODE) based on floral science and technology literature was completed, and semi-automatic indexing based on ontology knowledge base was achieved (see Figure 1).

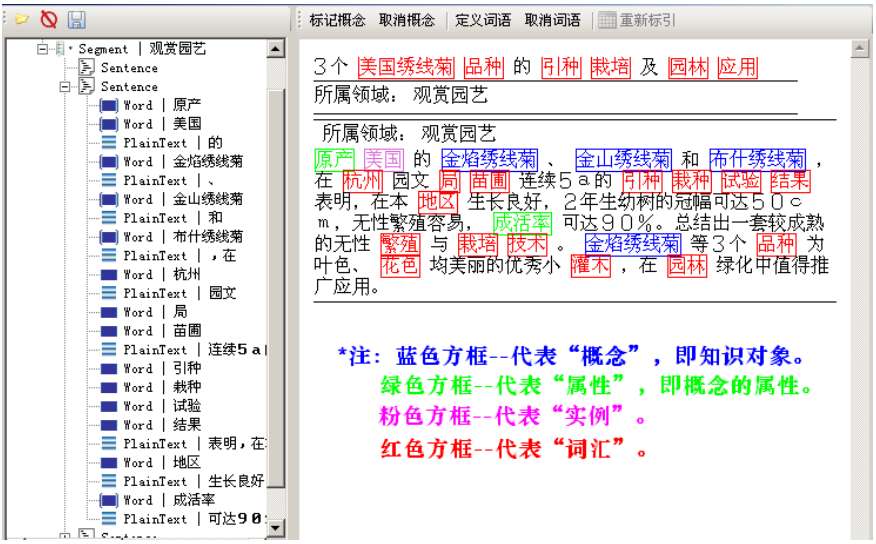


Fig 1 Example image of semi-automatic semantic indexing for Chinese scientific and technological literature (system automatically generates pages)

In 2011, the Ministry of Science and Technology supported the project "Construction and Application Demonstration of Knowledge Organization System for Foreign Language Science and Technology Literature Information" (2011BAH10B00, also known as the "Super Science and Technology Vocabulary Project") led by the National Science and Technology Library and Literature Center (NSTL). The construction goal of the project is to adopt internationally advanced knowledge organization technologies

and methods, draw on the existing knowledge organization system construction achievements and application experience at home and abroad, Strive to basically establish a scientific and technological knowledge organization system for computer applications during the 12th Five Year Plan period, provide support for the organization and utilization of China's massive foreign scientific and technological literature information, achieve effective organization, deep disclosure, and knowledge correlation of national scientific and technological literature information strategic resources, provide knowledge retrieval services, and promote knowledge discovery, knowledge mining, and knowledge computing application demonstration based on national scientific and technological literature information strategic resources, Overall enhance the knowledge service capabilities of China's scientific and technological literature and information institutions. Its specific goals include: building a foreign language scientific and technological knowledge organization system with domain ontology as the goal; and establish a sustainable development mechanism for the national scientific and technological knowledge organization system.

2 The Function of Ontology in Documentation

2.1 The Function of Ontology in Library and Information Science

Specifically, in the fields of information retrieval and knowledge organization in library and information science, ontology has the following functions and is applicable to all standard literature:

(1) Reflect the semantic mapping relationship and semantic limitations of vocabulary.

If we only search some title based on the search terms entered by the user, it will definitely lead to the result "missed detection". The search term entered by the user is related to the user's own knowledge background, search ability, and search experience, and may only be one of several synonyms, homoionyms, or related terms of a certain concept. Researchers can use ontology to standardize the concept set and automatically map search terms to their synonyms, synonyms, and related words, using a set of standardized concepts for retrieval.

(2) Structured organization of information resources based on a certain knowledge organization system.

For a long time, libraries have adopted classification and thematic methods to catalog their collections in order to establish information collections based on a knowledge organization system. In the network environment, using the idea of ontology to reconstruct online information resources can establish an ontology based information portal or knowledge portal; and use composite (integrated) ontology to index information sets from different perspectives.

(3) Represents the link between information content and the knowledge organization system.

Ontology can be linked to information systems, enabling users to more conveniently use ontology to understand specific concepts and link related concepts and resources during the process of using information. The linking method can be static (that is, the relevant links are embedded in the information unit beforehand and cannot be modified), or dynamic (that is, the system extracts vocabulary and links the corresponding ontology when needed). This type of application is generally used in professional fields, so the work of semantic analysis and ontology selection is relatively clear and concise.

(4) Utilize multiple patterns to represent and understand information sets.

By utilizing ontology, information can be visualized and organized. The retrieval results can be organized and displayed according to the classification system of the ontology. The collection of literature can be displayed according to a certain semantic relationship, for example, according to the relationship of "performing certain operations (identification, reproduction, cultivation, breeding, application, plant protection, quarantine...) on a certain object (a certain flower) in a certain environment (time, location, climate conditions, soil conditions)", displaying a specific type of flower or a certain type of flower, Literature search and result display of a certain cultivation measure under what cultivation conditions. Geographic information can also be combined with maps for display.

2.2 The Function of Ontology in Standard Literature

Standard literature belongs to the category of scientific and technological literature, and is a special type of literature. It is different from general scientific and technological literature due to its fast update and replacement cycle (standards have standard age), concentrated technical content, multiple data formulas and charts, and different types of standard text with their own fixed format and different formats. Based on the above, the application of ontology theory in standard literature services can play the following functions:

(1) Improve the retrieval efficiency (recall, precision) of standard literature databases for users.

(2) Establish ontology based knowledge linkages in the standard literature database, and provide users with composite retrieval, packaging services, and thematic services based on this.

(3) Applied in reference consulting services (expert systems), it automatically provides answers to users based on the accumulation and update of ontology knowledge base, logical relationship judgment between concepts in the ontology. And under the self-learning mechanism and automatic verification mechanism, it continuously accumulates and updates, making the answers continuously perfect and improving in quality.

3 Practical Research

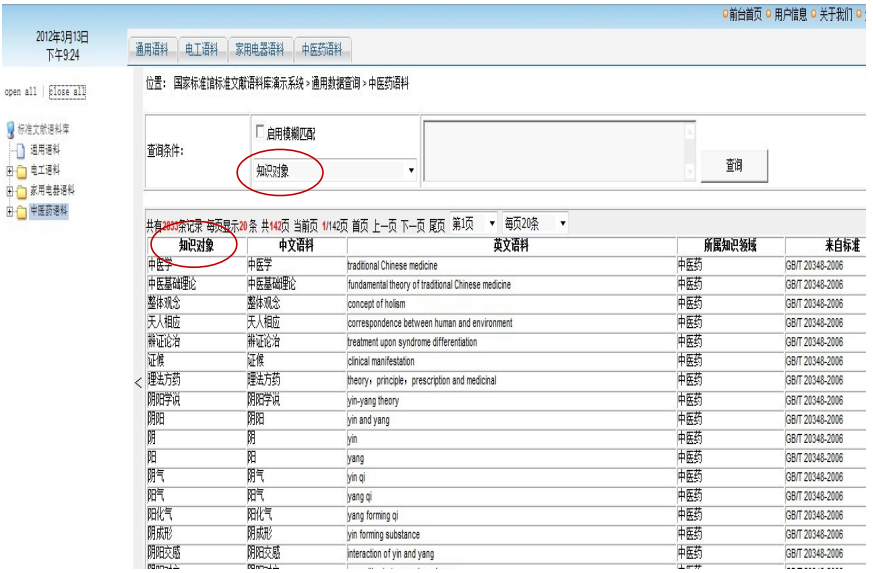
3.1 Construction of Standard Literature Corpus

Corpus, in linguistics, refers to a large amount of text, usually organized and with established formats and markers. Corpus refers to a data warehouse that stores raw language materials. A general corpus has a wide range of sources for processing its corpus, including books, dictionaries, monographs, etymologies, etc. The processing of standard literature corpus mainly focuses on standard literature and technical regulations. A knowledge base is a special tool used for knowledge management to facilitate the collection, organization, and extraction of knowledge related to the field of knowledge. The knowledge object in the knowledge base is a collection of professional domain knowledge required to solve problems, including basic facts, rules, and other relevant information.

The construction of corpora, knowledge bases, and knowledge organization systems (KOS) is a hot topic of knowledge management and services. The fundamental purpose of corpora is to provide the foundation and support for realizing functions and modules such as semantic indexing, machine translation, knowledge linkage, data mining, and intelligent retrieval.

The corpus prototype system developed in this study achieves professional navigation based on a simple knowledge organization system (primary ontology) (see Figure 2).

And in the data structure, the knowledge object field has been set up, which is used to establish a mapping table foundation for building a complex ontology based knowledge base, including "knowledge object" - "Chinese corpus" - "English corpus" - "XX language corpus". This setting is used to support the expansion of various functions such as machine translation and automatic indexing.



知识对象	中文语料	英文语料	所属知识领域	来自标准
中医学	中医学	traditional Chinese medicine	中医药	GB/T 20348-2006
中医基础理论	中医基础理论	fundamental theory of traditional Chinese medicine	中医药	GB/T 20348-2006
整体观念	整体观念	concept of holism	中医药	GB/T 20348-2006
天人相应	天人相应	correspondence between human and environment	中医药	GB/T 20348-2006
辨证论治	辨证论治	treatment upon syndrome differentiation	中医药	GB/T 20348-2006
证候	证候	clinical manifestation	中医药	GB/T 20348-2006
理法方药	理法方药	theory，principle，prescription and medicinal	中医药	GB/T 20348-2006
阴阳学说	阴阳学说	yin-yang theory	中医药	GB/T 20348-2006
阴阳	阴阳	yin and yang	中医药	GB/T 20348-2006
阴	阴	yin	中医药	GB/T 20348-2006
阳	阳	yang	中医药	GB/T 20348-2006
阴气	阴气	yin qi	中医药	GB/T 20348-2006
阳气	阳气	yang qi	中医药	GB/T 20348-2006
阳化气	阳化气	yang forming qi	中医药	GB/T 20348-2006
阴成形	阴成形	yin forming substance	中医药	GB/T 20348-2006
阴阳交感	阴阳交感	interaction of yin and yang	中医药	GB/T 20348-2006

Fig 2 Prototype system page for browsing corpus by profession

3.3 Topic Integration Based on Primary Ontology

Establish a primary ontology for home appliance products based on existing classification systems [such as standard classification, International Convention for Harmonized Commodity Description and Coding System (Harmonized System, HS) ,etc.], and classify and summarize relevant standard literature according to the established home appliance product ontology (see Figure 3). The advantage of thematic integration established in this way is that it is easy to query and browse, and conforms to the reading habits of general readers.

Fig 3 Home appliance product thematic integration product classification page

3.4 Research on Key Standards for Correlation between Standard Technical Indicators and Basic Product Information Knowledge

From 2013 to 2017, relying on the public welfare special project of the quality inspection industry, "Research on Key Standards for the Correlation between Standard Technical Indicators and Basic Product Information Knowledge", we mainly developed methods for extracting, analyzing, and statistical analysis of standard technical indicators based on basic product information. We studied the distribution patterns of general standard technical indicator systems, universal analysis methods and models, and established comparison rules and metadata processing specifications; conduct research on tracking technology for product standard technical indicators and establish an indicator tracking model. Afterwards, on the basis of establishing and improving the standard technical indicator system and product standard technical indicator tracking model, taking renewable energy related products (such as solar photovoltaic power generation products) enterprises and product standard technical indicator system as examples, the research and construction of the ontology based conceptual model and product standard indicator data statistical analysis

model related to knowledge elements such as products, standards, indicators and enterprise information were completed. We conducted business function testing and pilot work to verify and test the logic of the conceptual model and the practicality of the prototype system. The logical relationship between solar photovoltaic cells and technological standards and the lightweight ontology formed by enterprises is as follows:

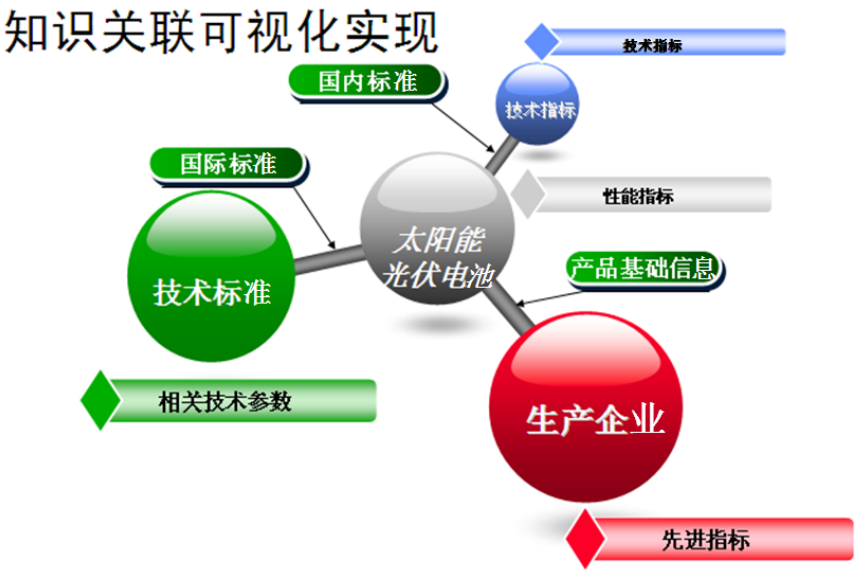


Fig 4 Logical Relationship Diagram between Solar Photovoltaic Cells, Standards and Enterprises

At present, the relevant achievements and technical indicator disclosure technology based on standard literature have been successfully applied for and officially published as the national standard GB/T 39872-2021 "Data Specification for Technical Indicator Disclosure of Standard Literature" in 2021.

4 Conclusion

Based on the above, in relevant research, the "Quality Analysis of Standardized Terminology and Corpus Application Research" focuses on the construction of ontology based standard literature knowledge base. The "Research on Knowledge Linkage of Technical Standards" focuses on the research and development of key technologies for ontology based knowledge linkage. The "Research on Key Standards for the Correlation between Standard Technical Indicators and Basic Product Information Knowledge" focuses on the disclosure of ontology-based technical standard content indicators and the study of knowledge correlation, and reflects the display and application of knowledge correlation between the four major

categories of knowledge elements: "enterprise—product—standard—technical indicators".

Based on the needs of enterprises and the market, the above research will continue to promote and support the translation of standard literature, service-oriented knowledge linkage of standard literature, and the construction of standard literature knowledge base.

The construction of ontology based knowledge organization system and corresponding key technology research and development will greatly (1) promote the improvement of knowledge service capabilities of standard literature information systems; (2) Promote the transformation of standard literature retrieval from traditional retrieval and query models to intelligent retrieval models with functions such as knowledge navigation, automatic topic clustering, content related disclosure, natural language retrieval, and bilingual (multilingual) query; (3) Promote the transformation of standard literature service work from traditional provision of standard literature information to knowledge service direction such as knowledge discovery of scientific and technological object entities, standardization hotspot monitoring, and trend analysis of standard formulation. (4) Promote the integration and correlation of standard literature and technological literature, leverage collaborative efforts, and better provide integrated and diversified knowledge services for users.

Reference

1. Gruber, T.R. A Translation Approach to Portable Ontology Specification [J]. Knowledge Acquisition, 1993(5): 199~220
2. Borst, P. and Akkermans, H. An Ontology Approach to Product Disassembly [P]. EKAW 1997, Sant Feliu de Gu5xols, Spain, October: 15~19
3. Uschold, Mike and Michael Gruninger. Ontologies: Principles, Methods and Applications [J]. Knowledge Engineering Review. Vol.11, 1996 (2): 93~136
4. Musen, M.A. Dimensions of knowledge sharing and reuse [J]. Computers and Biomedical Research, 1992(25): 435~467
5. UNSPSC. United Nations Standard Products and Services Code [EB/OL]. http://www.unspsc.org/（2003-05-01）
6. McGuinness, D.L. and Wright, J. Conceptual Modeling for Configuration: A description Logic-based Approach [J]. Artificial Intelligence for Engineering Design, Analysis, and Manufacturing - special issue on Configuration. Vol.12, 1998(4): 333~344
7. McGuinness, D.L., Fikes, R., Rice, J. and Wilder, S. (2000). An Environment for Merging and Testing Large Ontologies. Principles of Knowledge Representation and Reasoning [P]: Proceedings of the Seventh International Conference (KR2000). A. G. Cohn, F. Giunchiglia and B. Selman, editors. San Francisco, CA, Morgan Kaufmann Publishers
8. M. Ashburner et al. Gene ontology: Tool for the unification of biology. Nat. Genet., Vol.25, 2000(1): 25~29

9. Li Jing, Meng Xian xue, Su Xiao lu. Research on the Construction Method and Application of Domain Ontology [M], Beijing: China Agricultural Science and Technology Press, 2009.

Acknowledge

This work was supported by Projects of Fundamental Research Funds for the Central Institutes “Research on long-term preservation policy mechanism of standard scientific data”(252022Y-9466) and “Construction of the Standard System and Development and Application of Key Standards in the Digital Standard Library” (252023Y-10411).