Universe Scientific Publishing

ORIGINAL ARTICLE

Research and Application of Business-Driven Classification System for Atmospheric Environmental Data Resources

Jiaqi Huang*

Gefangjige Academies. E-mail: 1658099337@qq.com

Abstract: This article discusses the functional transformation of classification in knowledge management on the basis of summarizing the development of classification system, and analyzes the category characteristics and limitations of environmental industries under different classification systems. In this article, the idea and content of data classification and metadata service of the Federal Enterprise Architecture Framework (FEA Framework), which is a classification system with data sharing as the core under the background of big data transmission, are sorted out. The construction of comprehensive data collection and sharing platform for atmospheric environmental science is taken as an example to explore the business-driven scientific data sharing system. The result shows that with the transformation of knowledge carrier and dissemination mode, classification method has changed from knowledge structure, knowledge discovery to sharing. With the use principle of co-construction and sharing of network information, knowledge community and government information, the method has changed from traditional subject classification to business-oriented, which includes 11 categories to develop a metadata registry and full-text retrieval services to meet the characteristics of big data use.

Keywords: Atmospheric Environment; Big Data; Information Sharing

Introduction

Classification is an effective way to organize knowledge, and the commonly used classification methods are systematic classification and faceted classification. With the rapid development of computer and Internet technology, knowledge ontology has gone through the development stages of literature, informationization and digitalization, and accordingly, books and materials classification, network information classification, e-government classification, etc. have been produced. Classification adds new functions of knowledge discovery from single knowledge organization function. Classical classification systems, such as Chinese Library Classification and Dewey Decimal Library Classification, are developing towards automation and informationization, realizing the transformation from traditional manual paper classification to automatic and networked information classification. In this process, user-driven classification began to appear. With the rise and development of e-government, directory service based on metadata standard has become the technical standard of information sharing and reuse, and classification is the foundation of e-government directory system.

1. Classification methods and dif-

doi: 10.18686/pes.v2i2.1334

ferent classification systems

1.1 Classification methods

Traditional classification methods commonly include system beauty classification, faceted classification and mixed classification. System classification, also known as line classification and hierarchical classification, is usually based on scientific classification, which divides categories according to some characteristics or attributes of classified objects, refines categories step by step, and develops classification segments one after another to form a hierarchical tree-like classification system. In the category design of major search engines, the traditional classification idea is adopted. Its advantage is that the literature is concentrated according to disciplines and specialties, and the categories are developed systematically; Secondly, by adopting the method of ranking, we can clearly express the subordinate relationship and juxtaposition relationship among various items in the category system. The main disadvantage is that the class table is a static structure, which can not be changed at any time according to needs. Various complex topics and small special topics reveal in detail that the basic principle of restricted facet classification is to choose the essential characteristics or attributes of classified objects as facets; The same classification basis should be adopted for the same "face"; Categories in different "faces" do not cross each other and cannot be repeated. Classical faceted PMEST refers to personality, matter, energy, space and time. Faceted classification meets the needs of organization and search in specific fields in digital environment, and is designed to meet the professional and personalized faceted classification table, which is mainly used for website information structure, e-commerce product catalogue, enterprise content organization tools and post-control vocabulary for improving search efficiency. Network information is interactive, multi-dimensional, diverse, high-frequency and massive, so the application of traditional classification is limited.

The popular classification, also known as social label system, is a planar non-hierarchical structured label classification system defined by the spontaneity of the masses, which is one of the typical products of Web2.0. It allows users to label various types of network resources spontaneously in the form of metadata, and realizes resource sharing through tags, helping community users to search, browse, organize, share and innovate knowledge. Therefore, it is widely used in online communities at home and abroad. It has the characteristics of inclusiveness, flexibility, dynamics and user-centered information organization. At present, the mainstream network information classification mainly includes: a strict hierarchical classification system that integrates the theme-divided catalogues marked by the public; planarization classification system of mass classification based on theme method; user-driven non-strict hierarchical mass classification ontology classification system and automatically derived non-strict hierarchical mass classification ontology classification system, etc.

1.2 Classification system

Chinese Classified Thesaurus based on CLC (Web Version 8 is referred to as "Sub-list") provides thesaurus browsing, retrieval and data downloading services through the Internet, provides B/S mode library business support, and can interface with Online Public Access Catalog,OPAC). The further standardized processing of Sub-table needs the latest methods and technologies of Chinese ontology information processing, such as automatic word segmentation tagging, new word discovery, information extraction and automatic clustering. However, for a long time, the revision and maintenance of the Sub-table is still inseparable from the manual classification of environmental science in the Chinese Library Classification (**Figure 1**), which is the basic theory of environmental science, society and environment, environmental protection management, disaster and its prevention, environmental pollution and its prevention, waste treatment and comprehensive utilization, environmental quality evaluation and environmental monitoring, and safety science. From 1999 to 2010, the environmental classification of Chinese Library Classification was greatly adjusted, and 30 secondary categories were added. However, there are three same topics in the application of professional classification: Chinese Library Classification is a system classification, and the system classification belongs to a pre-listed category. The inherent defect is that it can not list the known topics exhaustively, and it can not absorb

the emerging new concepts in time; Second, each edition of Chinese Library Classification has a long publishing cycle, averaging about 10 years, which is incompatible with the rapid development of modern environmental science. Third, modern science has the characteristics of highly detailed and comprehensive development, which is more prominent in professional classification.

1.3 Classification of environmental information

In GBT7027-2002 "Basic Principles and Methods of Information Classification and Coding" defines that information is the truth and related statements of all meaningful concrete or abstract things or concepts, which are expressed through data, messages and their progress details; Clarify the relationship between information and data, that is, in the field of information classification and coding, the expression form of information is data. A correct understanding of various information concepts depends on information classification, and the unanimous expression of various information depends on information coding. The definition of information classification is to distinguish and classify information according to certain principles and methods according to the attributes or characteristics of information content, and to establish a certain classification system and arrangement order. Information classification includes three methods (line, area and hybrid) and five principles (scientificity, systematicness, extensibility, compatibility and comprehensive practicability). HJT417-2007 Classification and Code of Environmental Information was promulgated and implemented in 2007. It adopts a hybrid classification based on line classification supplemented by area classification for environmental information, and codes at category level, up to four levels.

There are 10 first-class categories in this standard, including environmental quality information, ecological environment, pollution sources, environmental management business, environmental science and its management, environmental protection industry, environmental government affairs management, environmental policies and regulations standards, environmental protection related letters, and other environmental information, etc. The atmospheric environmental quality data is the third-class beautiful item (**Figure 2**), which includes gaseous pollutant data, precipitation data, particulate matter data, ozone and greenhouse gas data, and other atmospheric environmental quality data, etc. Compared with the precise operational data produced by the joint observation of sky-earth stereo observation network, the classification and category arrangement of atmosphere in this standard is too macroscopic, which does not meet the classification standard requirements of comprehensive data of atmospheric environmental science. Therefore, according to the effective scope and capacity of environmental information classification and coding, the author will determine the specific classification method and structure of the comprehensive data collection and sharing platform for atmospheric environmental science. Based on the scientific data produced by the special scientific and technological work in the field of resources and environment from 1999 to 2012, Hongzhi Wang classifies the scientific data in the field of resources and environment by adopting the classification system of first-level and second-level disciplines and third-level keywords, with a total of 14 first-level classifications, 67 second-level classifications and 435 third-level classifications It can be seen from **Table 1** that the advantage of its classification is that it takes environmental media and characteristic pollutants as the three-level classification, which is easy to understand; The disadvantage is that there are too few characteristic pollutants.

2. Classification of FEA framework data

The federal enterprise architecture framework (FEA framework for short) is a methodology and integration tool for the top-level design of American e-government, which has a history of more than 10 years. the latest version was released in 2013, aiming at promoting information sharing, interoperability and common business process sharing development between various departments of the federal government and other government entities. The core of FEA framework is CRM, which is composed of six sub-models, including performance reference model, business reference model, data reference model, application reference model,

infrastructure reference model and security reference model. The FEA framework helps the federal government to realize the top-down e-government design and overall management, and guides the federal agencies to transform the government's strategic goals into institutional goals, and then refine them into concrete and implementable e-government projects. Therefore, it has also been absorbed by many countries and reflected in the reference model by the classification idea of FEA framework. It represents the main elements of the e-government system from six angles of performance, business, data, application, infrastructure and security. Except the data reference model, each reference model has 3~4 basic classification levels, which refines the specific content of the reference model layer by layer. For example, the business reference model is used to describe the main business flows and activities of the federal government by taking three definitions: business domain, business line and sub-function; The reference model is applied to classify the business and performance objectives of supporting government from three levels: system, application component and interface.

The data reference model (DRM) is oriented to business activities, focusing on the challenges of information interoperability and sharing between organizations and departments, and is the basis for realizing the trust description, discovery, management and sharing among the entire federal government, and regards government data as the national asset management principle. DRM is used to identify what data the federal government owns and how to share this data according to business tasks. DRM adopts three levels of classification, the top level is domain (4 items), followed by theme (22 items) and special topic (144 items). It should be noted that DRM classification standards are not fixed. On the contrary, it is flexible and extensible. With the change of the business model of the federal government, new topics and topics can be added. It also allows organizations to further decompose topics into organization-specific business processes according to needs. DRM points out that organizations and organizations participating in the COI of the US federal government classify data, define, compile and publish classified metadata that potential users can see and access, discover information, and then achieve the goal of data service. According to the business support characteristics and application scope of data, metadata is divided into six categories from bottom to top, namely, data concept, data exchange, data assets, topics, themes and domains. It should be made clear that data assets are not DRM classification standards, but they can be classified according to DRM taxonomy. In addition, data users can subscribe to topics published in the data registry to enhance data discovery. Once the data registry is shared, these classifications will become effective tools for data discovery and promote data sharing and reuse.

3. Conclusion

With the changes of knowledge carrier and dissemination mode, the function of classification has transformed from knowledge organization to knowledge discovery. Knowledge community and government information sharing, the classification method is based on business analysis from traditional disciplines with the use principle of network information. Eleven categories of atmospheric environment comprehensive data classification systems are established, and metadata registration is formulated.

References

1. Zhao Z. Analysis and thinking of atmospheric environmental monitoring (in Chinese). Resources Economization & Environmental Protection 2019; (09): 76. Doi: 10.16317/j.cnki.12-1377/x.2019.09.067.
2. Yang D, Zhu S, Wang Z. Application of big data analysis technology in atmospheric environmental monitoring (in Chinese). Technology Wind 2019; (27): 126. doi: 10.19392/j.cnki.1671-7341.201927116
3. Tu J. Discussing the application of big data analysis technology in atmospheric environmental monitoring (in Chinese). Low Carbon World 2019; 9(08): 1-2.