doi:10.3772/j.issn.1006-6748.2019.02.013

Method and application of knowledge organization system for S&T literature①

Zeng Wen(曾 文)*, Li Hui②**, Fu Hong**, Wang Zhongjun***, Yu Wei***, Xu Zhen*
(*Institute of Scientific and Technical Information of China, Beijing 100038, P. R. China)
(**Beijing Institute of Science and Technology Information/Beijing Science and Technology Strategy Decision-making Consulting Center, Beijing 100089, P. R. China)
(***North Institute for Scientific & Technical Information, Beijing 100089, P. R. China)

Abstract

In the big data environment, the construction of massive S&T literature data resources needs intelligent technical assistance. With a focus on comparing the domestic and foreign knowledge organization systems and their applications, this article analyzes and summarizes the gaps in related researches and applications at home and abroad. A knowledge organization system framework for S&T literature data resources is presented in the article. Starting from the basic element of knowledge organization system, it also proposes and designs terminology-based analysis methods and technologies for S&T literature. Based on this framework, it proposes ideas and develops corresponding software tool to carry out relevant experiments. It gives an overview of theories and technologies method for future research.

Key words: scientific and technological (S&T), knowledge organization, term, intelligence analysis

0 Introduction

With the advent of big data era, the S&T literature of different types and sources increase in scale day by day. The key problem is how to obtain valuable information from these complex and massive S&T literature. The organization tools of S&T literature, such as S&T thesaurus, can be used for organization of S&T literature data. The knowledge organization system based on the thesaurus is one of the important ways for effective organization and application of massive S&T literature. The knowledge organization system is the term and symbol system that defines and describes the objective material world information and interrelationship, which rationally describes and organizes various S&T literature according to the knowledge content and structure, and realizes knowledge navigation and correlation analysis. It is mainly used to explain certain concept, concept attributes and concept relationship, and the knowledge organization system is a bridge between user's information needs and information resources, and also the core of information resource construction. The application of the knowledge organization system involves the following aspects: (1) Control the synonym and near-synonym through the expression of certain concepts, concept attribute and relationship between concepts, construct the knowledge according to certain structures, and support semantic understanding, specification and navigation towards information resource retrieval through the description and organization of information data and the relationship between existing concept knowledge[1,2]. (2) Improve the efficiency of retrieval service and promote the sharing and exchange of domain expert knowledge through the established concept or knowledge, concept map, thematic map and ontology[3,4]. (3) Conduct concept analysis, classification, and indexing of information according to the concept and relationship of the knowledge organization system, which realizes semantic-based retrieval, and is convenient for comprehensively and thoroughly revealing the information, relationship between information and its position in the whole knowledge organization system[5-8]. The concept of knowledge organization system proposed in this article is not only the construction idea of knowledge organization system which only operates knowledge elements, but also aims at knowledge innovation and service.

① Supported by the National Social Science Fund of China (No. 18BTQ054).
② To whom correspondence should be addressed. E-mail: lisa-lh@126.com
Received on Nov. 19, 2018

1 Related work

1.1 Comparison of research status at home and abroad

The knowledge organization system aims to maximize the description of resources and contents based on analysis of existing resources and users' needs, so as to provide data resources that meet user needs. This section investigates and analyzes the knowledge organization systems and their applications at home and abroad, as shown in Table 1.

Table 1 Main knowledge organization systems and applications at home and abroad

Knowledge organization system type	Application example	Implementation method	Application field
Thesaurus, subject headings	Web of service	Subject headings	Cross-database cross-platform citation/ subject retrieval, browsing and result analysis
	SOSIG HILT UMLS	Interoperability between thesaurus and classification thesaurus	
	NSTL		Cross-domain retrieval and browsing
Classification method, subject method	CNKI Baidu Baike	Classification-subject integration self-edited classification method, subject method	Discipline classification, subject retrieval knowledge classification, topic clustering retrieval
Ontology	GoPubMed AOS	Thesaurus/semantic association technology Top-level integration domain ontology	Multi-angle integration analysis, statistics and display of retrieved results
Knowledge mapping	Google	Knowledge entity database	Retrieval of text and entity links

Research results show that most existing literature-based knowledge organization systems in China, except for CNKI and Baidu, transform existing vocabularies or thesaurus. Such knowledge organization systems are limited in application. In China, there are few cases of successful application of knowledge organization systems. Even the relatively successful knowledge organization systems of CNKI and Baidu have not reached the level of foreign knowledge organization systems. How to deeply reveal the knowledge of information resources from multiple perspectives, provide a variety of information resource navigation and integrated retrieval services, and allow users to obtain knowledge in real time? The main approaches are: (1) The status of information resources is an important factor affecting the design and application of knowledge organization system. As information resources increase in quantity and type, higher requirements for the knowledge organization system are raised. Knowledge organization system should be designed and implemented according to actual situation of information resources and application needs. A comprehensive knowledge organization system is needed for the organization of comprehensive information resources. For organization of professional information resources, a professional knowledge organization system should be designed for certain field or discipline. The general knowledge organization system of comprehensive information resources is usually complete and extensive, while the professional knowledge organization system is refined having discipline characteristics. The design of knowledge organization system featuring huge resources, rich carrier types and rich disciplines or specialties, must be refined and diverse, so as to realize the multi-angle revelation, effectively integrated organization and retrieval of information resources. (2) Asymmetry between the knowledge organization specialization and data quality. As important parts of the knowledge organization system, vocabulary, thesaurus and subject headings, etc. are important tools for information organization and service. Thesaurus as a professional tool of library information is undoubtedly authoritative and professional, serving as one of the main tools supporting information organization and service. The famous foreign searching institutions also need thesaurus to support knowledge service, but another important factor of information service is the basic indexing of S&T information resource. Foreign S&T literature is strictly indexed according to the standard of thesaurus, so their information retrieval and knowledge services are more accurate. In China, no

institutions and personnel engage in similar basic work. Therefore, although there are many thesauri in various fields and scales in China, they are not better used for document retrieval and multi-domain knowledge navigation services. At present, the application of the knowledge organization system in China features high recall ratio, low accuracy rate and high operation cost basically. 3) Lack of practicable standards and norms for knowledge organization semantic interoperability. Referring to the capability of understanding the data shared between systems at the well-defined conceptual level, semantic interoperability is one of the main issues of information services in the network environment. Semantic interoperability needs unified norms. For example, concept representation, synonymous concept merging, semantic normalization, ambiguous concept processing, etc., are subject to certain norms and standards. As the formulation and improvement of domestic interoperability standards are very immature, the integration of vocabularies and classification tables is basically separated, with poor sharing property. Only by adhering to the unified norms and standards, realizing the synonymous concept merging and ambiguous concept processing through machine-assisted manual intervention, generating the normalized concept and unique identifiers, inheriting and normalizing the important codes of source table, and building multi-level term types and structured semantic nets, it can support the efficient and unambiguous data interaction between computer systems containing different vocabularies. The basic mechanism for semantic interoperability of information systems is as follows: conceptually organize data, define the synonymous concepts from each knowledge organization system, provide the semantically standardized names and unique concept identifiers, and make the synonymous concepts at a certain abstract level be in line with the same standardized name and code.

1.2 Key problems for construction of knowledge organization system

1.2.1 Unified representation of knowledge organization

The knowledge organization system needs unified representation forms, specification and technical interface to integrate other industry classification methods, subject headings and be compatible with existing subject headings of multiple formats. Traditional subject headings are an important knowledge organization tool for information resource management. As the demand for content-based information processing grows, subject headings as a knowledge system has become an important supporting tool for visual analysis and evolution analysis of concepts. The professional development of subject headings brings both convenience and problems to the management of domain information resources. For example, interdisciplinary research is usually active. In the keywords of a frontier paper, there are many terms across multiple fields, which are hardly covered by a subject heading concerning a certain field. In addition, subject headings compiled by various industries are saved in different ways, which brings a lot of inconvenience to the construction of the knowledge organization system. The subject heading compilers (domain experts) formulate subject headings from their respective perspectives, while users (information technology experts) of subject heading hope to have access to various subject headings of different fields through a unified software interface. Therefore, cross-domain and multi-source compatibility and integration are the important issues to be addressed for the knowledge organization system represented by subject headings. Therefore, the unified representation form, specification and technical interface of the knowledge organization system must be established, so as to integrate other industry classification methods and subject headings and be compatible with existing subject headings with multiple formats.

1.2.2 Semantic integration of knowledge organization

Integration of knowledge organization is not a merger of simple vocabularies, but an integration of words, concepts and relationships in different knowledge organization systems. Semantic conflicts between vocabularies shall be eliminated to form the consistent integrated vocabulary. In the process of semantic integration, in addition to the consistency processing of the structure layer and the concept layer, the deep semantic association between terms needs to be discovered through the mining technology of information resources: (1) Synonym and polyseme integration. The synonyms and polysemes in the cross-domain and multi-source vocabularies can be integrated at the glyph level. (2) Concept mapping. Multiple relationships between two subject headings such as complete equality and incomplete equality (most equal, partially equal) can be addressed through concept mapping. (3) Concept integration. As affiliation relationship has different meanings in different subject headings, the focus of concept integration is the discovery of concept affiliation relationship. (4) Relationship integration. The integration of vocabulary is not only the mapping integration of the same concepts, but also needs to discover new association between cross-disciplinary thesauri through the mining of data.

1.2.3 Analysis of user's behavior data

The purpose of analyzing and utilizing user's behavior information is to support the knowledge organization system service. The user behavior information is generated in the process of interaction between the user and the knowledge organization system, and the user's behavior rules and patterns are analyzed and summarized to achieve quick feedback of knowledge service. At present, the existing user's behavior analyses or researches mainly focus on the methods applicable after the knowledge organization system construction, and the scope is limited to the user's behavior influence factors, while the initiative of knowledge service is insufficient. This article holds that the user's behavior data analysis is a process of analyzing the accumulated user's behavior, but also a process of real-time tracking and acquisition of user's behavior and engagement in the knowledge mining. Therefore, the knowledge organization system should establish and cover user's behaviors, apply the data analysis techniques and methods, regard user's behavior analysis as "knowledge", and achieve the pertinent knowledge, so as to provide users with effective active information services.

2 Optimized framework of the knowledge organization system for S&T literature

The foreign researches and applications concerning the knowledge organization systems are increasingly diversified. The applications are not limited to the methods and technologies for the traditional knowledge organization systems, while richer and more perfect knowledge services can be delivered by use of the computer storage technology, retrieval technology, big data technology, etc. For example, interoperability methods and techniques for knowledge organization systems support the effective integration and retrieval of resources among different knowledge organization systems. Building and integration of digital environments, retrieval technology and content-based linking capacity, enable the seamless integration of high-quality information resources, information analysis tools and information management software. Searching of the documents with relevant semantic contents through semantic matching and reasoning will achieve the organization of related information meeting different needs, as well as the browsing and navigation of the S&T literature and knowledge based on semantic technology. Traditional classification scheme and vocabulary are not taken as the classification and organization tools of knowledge organization system, and new data classification and organization system based on the characteristics of data resources is established. In addition, it is more practical to flexibly establish the applicable knowledge organization system according to the user's need and information resource characteristics.

Therefore, this article proposes a knowledge organization system framework as shown in Fig. 1, which consists of four parts, namely the basic platform layer, data resource layer, business layer and service layer. Among them, the basic platform layer is the platform of knowledge organization system, namely the operating system and the database management system. The data resource layer is the data source of knowledge, and the basis of knowledge preparation, mainly involving various types of corpora for knowledge extraction, existing knowledge organization system resource and user's behavior resource. The business layer is the functional layer of knowledge organization system, which mainly functions to realize the knowledge processing, the integration, updating or extension of traditional knowledge organization system, and the analysis of user's behavior knowledge. The service layer mainly provides user-oriented services, mainly including information retrieval services, information knowledge organization and navigation, visualization and interface services of knowledge organization system, etc. Compared to the traditional framework of knowledge organization system, the frameworks shown in Fig. 1 has greater coupling and integration degree in data, function and service, and has the dual functions of the traditional integration and knowledge reconstruction of knowledge organization system. The user's behavior resource is increased for the data resource layer, and the analysis function of user's behavior knowledge is newly added for the business layer, make the information service of knowledge organization system more accordant with the users' need.

3 Realization of knowledge organization system for S&T literature

In this section, general idea and mechanism of the framework of knowledge organization system for S&T literature will be presented and the development process will be established. In the basic platform layer, operation system adopts Windows. In the data resource layer, real S&T literature and Chinese S&T vocabulary system edited by ISTIC is used. In the business layer, core functions, auxiliary or extended function are realized. In the service, retrieval, navigation and visualization of knowledge organization system are realized by developing software tools. The article will introduce emphatically term recognition and knowledge

extraction in the business layer and software tools for S&T literature in the service layer.

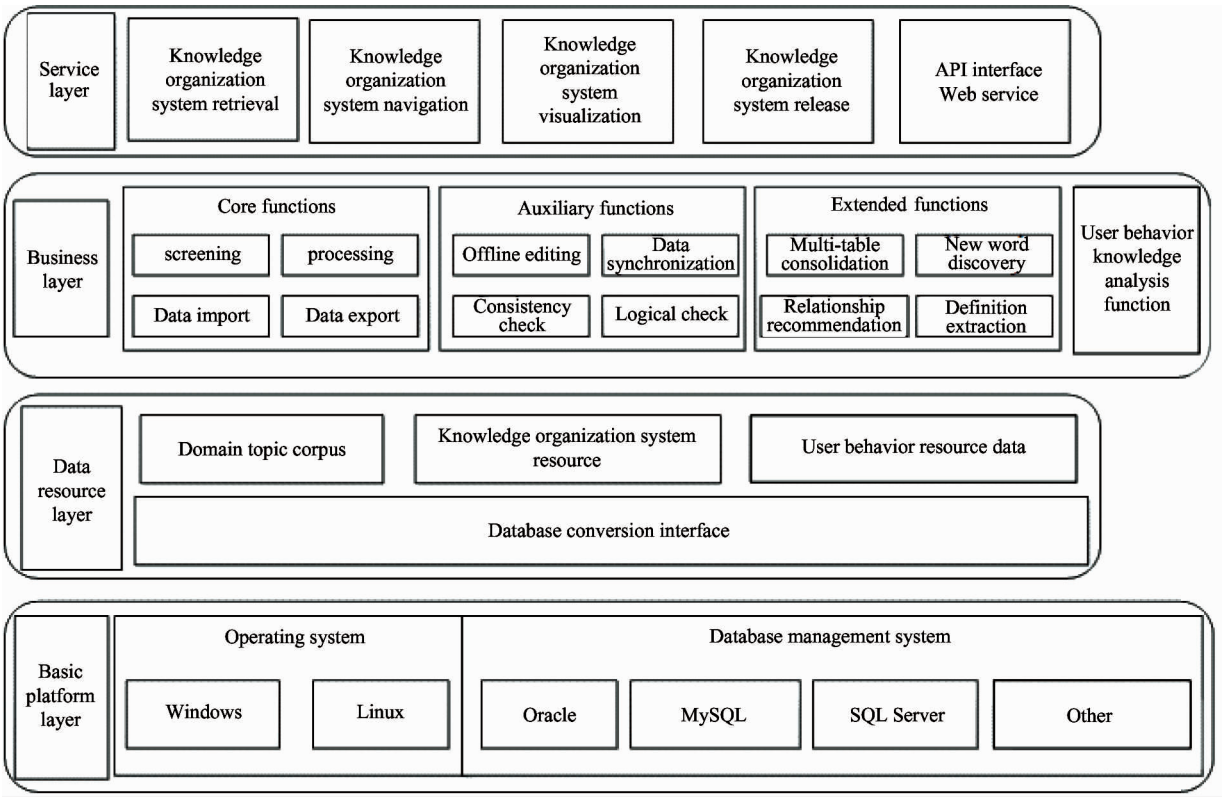


Fig. 1 Basic framework of knowledge organization system for S&T

3.1 Term recognition and knowledge extraction in the business layer of knowledge organization system

3.1.1 Term recognition

There is no uniform and standard definition of term, but many scholars have put forward their own views on the concept of term. Term refers to the collection of words used to express concepts in the professional field, which is closely related to knowledge and concepts in the professional field. There are some differences between term and ordinary vocabulary. Firstly, the field of term circulation is limited. Generally speaking, it only circulates in a fixed number of fields. It is only used by people who study in this field, while the field of common vocabulary circulation is very wide and can be used by all people. Secondly, term generally only exists in this field. It has a high degree of circulation, and terms can be distinguished from common words in terms of meaning, morphology, syntax and grammar. Term recognition is an important research subject in the study of knowledge organization. Term is the linguistic expression of the concepts in professional knowledge, which are accumulated through incremental exploration and research in specific fields. Term recognition is also called term identification, which is the process of obtaining term representing professional concepts from texts in respective special fields. Manual recognition from corpus is very inefficient and insufficient to meet the needs of information analysis, especially in the era of big data. Therefore some methods are required to fulfill the task of term recognition. The concept of deep neural network is derived from the study of artificial neural network. It is an emerging learning method based on multilayer neural network, which aims to obtain high-level features that can better represent raw data through learning by multilayer structure. At present, study of deep neural network has become trendy and its application has been employed in some fields, but there is little research on term recognition. Therefore, this study in the article attempts to apply deep neural network to distinguish term and develop recognition method based on it, which can hopefully provide more accurate identification of terms from corpus at certain scale so as to help people engaged in information analysis to improve their work and lay foundation for the research.

As the tool of word analyzers usually divides words into one-character word, whereas the length of term varies a lot, the results of term recognition would be un-

der direct influence of original word division if the article uses the original corpus to train word vector and model of deep neural network, eg. AE. The article constructs a candidate term dictionary for the tool of word analyzers during dividing words. By analysis and summary of word formation properties about Chinese terms, one can find that most Chinese terms are binary or ternary terms. Terms are usually certain parts of speech such as nouns, verbs, adjectives, while conjectives, prepositions, interjections are hardly used as terms. Therefore, we consider part of speech and its combination to construct part of speech rules template of Chinese terms. Besides, even such parts of speech as nouns, verbs and adjectives that can become terms, there are some words that cannot become terms. Before matching terms, the article use rules template to filter words whose part of speech cannot be terms. The main method is to count word frequency of texts that have done word division and summed up high-frequency words whose part of speech are included in the language rules template but obviously cannot be terms, and finally make preliminary selection of candidate terms according to the language rules template.

The candidate terms are only preliminary ones, and many of them are not terms, for instance, some common word collations and meaningless collations. Therefore, the article selects terms from candidate terms by C-Value, which is a widely used method that combines rules with statistics. The calculation formula is:

$$\text{C-Value}(t) = \begin{cases} \text{Log}_2 |t| f(t) \\ \qquad \text{when } t \text{ is not a nested term} \\ \text{Log}_2 |t| \left(f(t) - 1/c(t) \sum_{i=1}^{c(t)} f(b_i)\right) \\ \qquad \text{when } t \text{ is a nested term} \end{cases} \tag{1}$$

where t represents a candidate term, $|t|$ represents the length of the term "t", whose value is the number of characters of t, $f(t)$ represents the word frequency, $c(t)$ represents the number of candidate terms for nested t, b_i represents candidate term of nested t, $f(b_i)$ represents word frequency of b_i. The higher the value of the C-Value is, the greater the possibility of candidate term becoming a term is. According to the value of C-Value, and candidate terms will be eliminated if its value is smaller than threshold value. The remained candidate terms can act as candidate terms for dictionary.

3.1.2 Knowledge extraction based term

Term has become an essential tool for people to carry out scientific and technical exchanges and academic research. This section takes term as a starting point to study and elaborate the methods and techniques of knowledge extraction. Current knowledge extraction methods are basically based on term in the vocabulary and thesaurus. Taking thesaurus as an example, thesaurus achieves the indexing and organization of information resources by establishing a glossary of terms, as well as the relationships, generation, subordination, division and reference of terms. Knowledge extraction is based on term structure, the relationship between terms. Relationship and structure of term become the key contents of knowledge extraction.

(1) Design of term structure

Vocabulary: core word, foundation word

Definition: each core vocabulary has one definition.

Relations: synonyms are symmetrical and transitive; genera are asymmetrical and transitive under the same subclass; correlation is symmetrical and generally non-transitive.

Classification mapping: multidimensional classification mapping can be set according to the related classification methods.

Attribute: it can be set according to the specific field and knowledge service requirements.

(2) Design of relationship between terms

Synonymy relationship: synonymy, antonymy.

Hierarchical relationship: genera, whole parts, cases.

Correlation relationship: interdependence.

(3) Method of term extraction

The system automatically calculates the word frequency of term in corpus, determines the accuracy of term extraction through expert verification, and re-imports the modified term frequency results into the system. In addition, the co-occurrence of terms in corpus is calculated so that when the user increases the correlation of the entries, the terms with more co-occurrences of the entries can be recommended to the user.

(4) Semantic relation expression and semantic promotion of terms

In knowledge extraction, the semantic description of term is equivalent to the description of knowledge, including term, term definition and the relationship between terms. However, for comprehensive, multi-domain knowledge extraction, it is necessary to establish unified standards and norms for knowledge organization semantics and interoperability implementation, so as to enhance and realize the ability of data sharing among systems to be understood at the level of fully defined domain concepts, specifically involving the expression mode of term concepts and synonymous concepts attribution, and with the semantic standardization, ambiguity concept processing and other technical means.

3.2 Developing of software tools in the service layer of knowledge organization system

The article proposes using the terms in knowledge organization system to analyze S&T literature. The important content or valuable content for users are identified from the S&T literature based on knowledge organization. Therefore, it is important that analysis of content should focus on the important sentences in the S&T literature. And the calculation of the weight about sentences is the basis of judging the importance of sentences. The article focuses on the calculation of multi-feature in a certain way. The method can obtain relatively important sentences according to the calculated values. The used features may include: frequency of a term, the similarity between the title and sentence, sentence position and so on. Based on the characteristics of S&T literature, this article uses frequency of a term and similarity between the title and sentence. The terms are distinguished by methods proposed in the previous work[9]. In addition, the article uses Chinese S&T vocabulary system edited by ISTIC to select the important and technical terms, strengthen the accuracy of judgment about the importance of the sentence in the S&T literature. The main steps include: 1) Calculation of term frequency in the Chinese S&T literature. 2) Calculation of similarity between sentence and title. 3) Calculation of technical strength.

In order to verify the method of analysis in the article, a software tool is developed to realize the proposed algorithm and method in the article. The main mechanism and flow chart of developing software tool is shown in Fig. 2. The analysis tool interface is shown in Fig. 3. The result of information analysis is shown in Fig. 4. which includes 政策标号 (serial number of a S&T policy), 标题 (the title of a S&T policy) and 句子 (sentences of a S&T policy). 重要度(importance value).

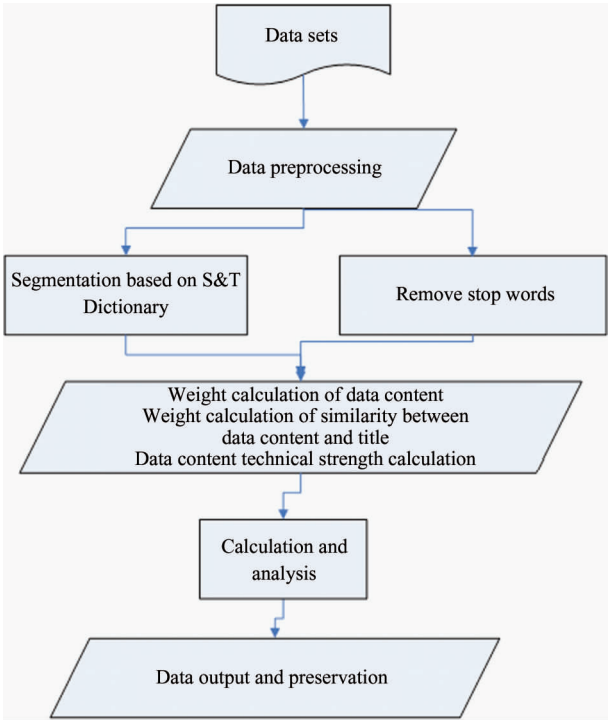


Fig. 2 The main mechanism of the developing software tool

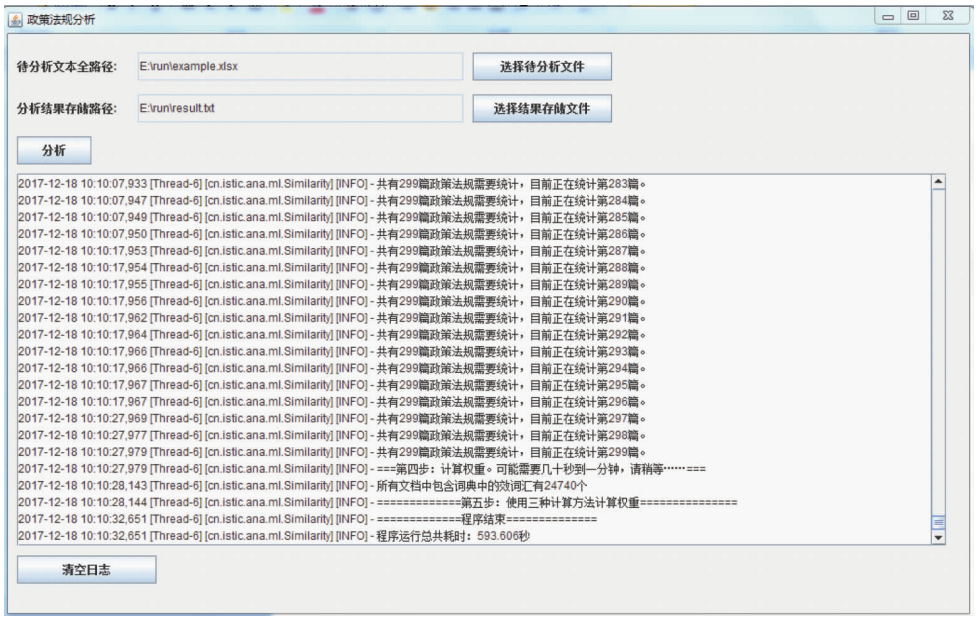
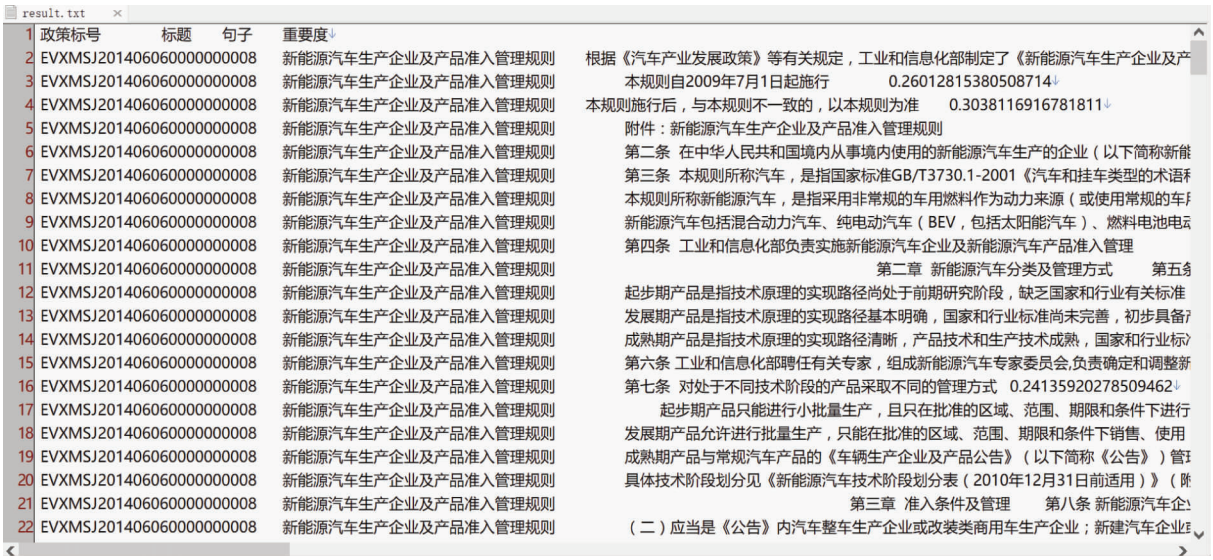


Fig. 3 The analysis tool interface shown in Fig. 2



result.txt

	政策标号	标题	句子	重要度	
2	EVXMSJ2014060600000000008	新能源汽车生产企业及产品准入管理规则	根据《汽车产业发展政策》等有关规定，工业和信息化部制定了《新能源汽车生产企业及产		
3	EVXMSJ2014060600000000008	新能源汽车生产企业及产品准入管理规则	本规则自2009年7月1日起施行	0.26012815380508714	
4	EVXMSJ2014060600000000008	新能源汽车生产企业及产品准入管理规则	本规则施行后，与本规则不一致的，以本规则为准	0.3038116916781811	
5	EVXMSJ2014060600000000008	新能源汽车生产企业及产品准入管理规则	附件：新能源汽车生产企业及产品准入管理规则		
6	EVXMSJ2014060600000000008	新能源汽车生产企业及产品准入管理规则	第二条 在中华人民共和国境内从事境内使用的新能源汽车生产的企业（以下简称新能		
7	EVXMSJ2014060600000000008	新能源汽车生产企业及产品准入管理规则	第三条 本规则所称汽车，是指国家标准GB/T3730.1-2001《汽车和挂车类型的术语和		
8	EVXMSJ2014060600000000008	新能源汽车生产企业及产品准入管理规则	本规则所称新能源汽车，是指采用非常规的车用燃料作为动力来源（或使用常规的车用		
9	EVXMSJ2014060600000000008	新能源汽车生产企业及产品准入管理规则	新能源汽车包括混合动力汽车、纯电动汽车（BEV，包括太阳能汽车）、燃料电池电动		
10	EVXMSJ2014060600000000008	新能源汽车生产企业及产品准入管理规则	第四条 工业和信息化部负责实施新能源汽车企业及新能源汽车产品准入管理		
11	EVXMSJ2014060600000000008	新能源汽车生产企业及产品准入管理规则	第二章 新能源汽车分类及管理方式	第五条	
12	EVXMSJ2014060600000000008	新能源汽车生产企业及产品准入管理规则	起步期产品是指技术原理的实现路径尚处于前期研究阶段，缺乏国家和行业有关标准		
13	EVXMSJ2014060600000000008	新能源汽车生产企业及产品准入管理规则	发展期产品是指技术原理的实现路径基本明确，国家和行业标准尚未完善，初步具备产		
14	EVXMSJ2014060600000000008	新能源汽车生产企业及产品准入管理规则	成熟期产品是指技术原理的实现路径清晰，产品技术和生产技术成熟，国家和行业标准		
15	EVXMSJ2014060600000000008	新能源汽车生产企业及产品准入管理规则	第六条 工业和信息化部聘任有关专家，组成新能源汽车专家委员会,负责确定和调整新		
16	EVXMSJ2014060600000000008	新能源汽车生产企业及产品准入管理规则	第七条 对处于不同技术阶段的产品采取不同的管理方式	0.24135920278509462	
17	EVXMSJ2014060600000000008	新能源汽车生产企业及产品准入管理规则	起步期产品只能进行小批量生产，且只在批准的区域、范围、期限和条件下进行		
18	EVXMSJ2014060600000000008	新能源汽车生产企业及产品准入管理规则	发展期产品允许进行批量生产，只能在批准的区域、范围、期限和条件下销售、使用		
19	EVXMSJ2014060600000000008	新能源汽车生产企业及产品准入管理规则	成熟期产品与常规汽车产品的《车辆生产企业及产品公告》（以下简称《公告》）管理		
20	EVXMSJ2014060600000000008	新能源汽车生产企业及产品准入管理规则	具体技术阶段划分见《新能源汽车技术阶段划分表（2010年12月31日前适用）》（附		
21	EVXMSJ2014060600000000008	新能源汽车生产企业及产品准入管理规则	第三章 准入条件及管理	第八条 新能源汽车企业	
22	EVXMSJ2014060600000000008	新能源汽车生产企业及产品准入管理规则	（二）应当是《公告》内汽车整车生产企业或改装类商用车生产企业；新建汽车企业		

Fig. 4 The results of analysis

4 Conclusions

The article puts forward ideas of knowledge organization system and carries out technological research and development for S&T literature. The article also proposes a basic method of knowledge organization system construction and it makes experiments to verify the idea and method on the real Chinese S&T corpus in the specific field. The experimental results show that the method proposed in this article is effective. In future research it will be improved to increase accuracy of analysis. Future research will mainly focus on the following aspects: Knowledge preparation; Knowledge acquisition, organization and integration; Semantic analysis of knowledge; Knowledge organization and service in line with user's behaviors[10,11].

Reference

[1] Zhao W, Liu X Q, Qiu B H. Construction of domain ontology in semantic web adaptive learning system [J]. *Journal of Jilin University* (*Information Science Edition*), 2008,(5):514-518 (In Chinese)

[2] Andres B, Ana B L, Ramon N S, et al. Research and technology organizations: how do they manage their knowledge? [J]. *International Journal of Entrepreneurship and Innovation Management*, 2007, 7(6):556-574

[3] Fausto G, Biswanath D, Vincenzo M. From knowledge organization to knowledge representation [J]. *Knowledge Organization*, 2014,41(1):44-56

[4] Wang J X. Knowledge organization tools and semantic interoperability method system[J]. *Digital Library Forum*. 2013, (5): 41-46 (In Chinese)

[5] Joo S, Choi I. Topic analysis of the research domain in knowledge organization: a latent Dirichlet allocation approach[J]. *Knowledge Organization*, 2018,45(2):170-183

[6] Wang H, Gu J, Su X N. Ontology-driven knowledge management system model and its application[J]. *Journal of Library Science in China*, 2013, 39 (2): 98-110 (In Chinese)

[7] Su X N. Knowledge Organization Theory and Method for Knowledge Services[M]. Beijing: Science Press, 2014. 18-19 (In Chinese)

[8] Zhang W X, Zhu Q H. Research and application of ontology-based semantic analysis process and method[J]. *Application Research of Computers*, 2011,28(3):961-964 (In Chinese)

[9] Zeng W. Term extraction and correlation analysis based on massive scientific and technical literature[J]. *International Journal of Computational Science and Engineering*, 2017,15(4):248-255

[10] Xiao H M, Hou Y. Problems and countermeasures of S&T information resources sharing under internet and environment [J]. *Technology Intelligence Engineering*, 2015, 1(6): 39-42. (In Chinese)

[11] Sun X P. Challenges of big data knowledge computing [J]. *Technology Intelligence Engineering*, 2015, 1(6): 43-50 (In Chinese)

Zeng Wen, born in 1973. She received her Ph. D degree in Shenyang Institute of Automation, Chinese Academy of Sciences in 2009. She also received her M. S. degree from the College of Information Science and Engineering at Northeastern University in 2003. Her research interests include information analysis method and technology, information theory and method.