# Challenges and Countermeasures for Knowledge Organization in Big Data Services: Postprint

**Authors:** Yunliang Zhang

**Date:** 2023-04-01T16:15:49+00:00

## Abstract

[目的/意义] The demand for big data services presents greater challenges to knowledge organization. By identifying, understanding, and analyzing these challenges, we can anticipate potential transformations in knowledge organization work and propose corresponding solutions.

[方法/过程] Focusing on the construction and application of knowledge organization systems, this study analyzes the challenges of knowledge organization from multiple perspectives based on real-world case studies of big data service projects, and proposes corresponding strategies.

[结果/结论] The challenges of knowledge organization in big data services can be categorized into four aspects: data explosion, literature assurance, integration, and application. To better address these challenges, we propose a series of knowledge organization frameworks for big data services, encompassing novel knowledge structures, multi-source updating strategies, and flexible application service models.

## Full Text

### Challenges and Countermeasures of Knowledge Organization in Big Data Services

**Zhang Yunliang**1,2 1. Institute of Scientific and Technical Information of China, Beijing 100038 2. Key Laboratory of Rich-media Digital Publishing Content Organization and Knowledge Service, Beijing 100038

#### Abstract

[**Purpose/Significance**] The demand for big data services presents greater challenges for knowledge organization work. By identifying, understanding, and analyzing these challenges, we can grasp potential changes in knowledge

organization and propose responsive methods. [**Method/Process**] Focusing on the construction and application of knowledge organization systems, this paper analyzes challenges from different perspectives of knowledge organization and proposes countermeasures based on real-world cases from big data service projects. [**Result/Conclusion**] The challenges of knowledge organization in big data services can be categorized into four aspects: data explosion, document assurance, integration, and application. To better address these challenges, we propose a series of knowledge organization frameworks for big data services, including new knowledge structures, multi-source updating strategies, and elastic application service models.

**Keywords:** knowledge organization; big data; knowledge service; challenge; countermeasure
**Classification Number:** G250
**DOI:** 10.13266/j.issn.0252-3116.2020.04.010

## Introduction

The concept of big data emerged early, but it only attracted widespread attention after analyst D. Laney defined the 3V characteristics in 2001 [1], profoundly influencing numerous industries including government [2], libraries [3], and intelligence agencies [4]. Big data services, which mobilize large amounts of distributed data and computing resources based on the characteristics of big data, can be divided into two types: big data query services and big data analysis services [5]. Both query and analysis services rely on organized knowledge, and knowledge organization processes such as processing, organizing, revealing, and controlling knowledge [6] contribute to service quality improvement [7]. The full realization of big data value requires broader internal associations within the data [8], which is precisely the core function of Knowledge Organization Systems (KOS). In this new environment, KOS itself also needs to evolve to cope with disciplinary and data changes while becoming more flexible and adaptable [1]. In big data service practices, knowledge organization systems such as ontologies, open linked data, and knowledge graphs have been widely applied.

Knowledge organization work for big data has received attention and promotion from institutions across different fields. Universities, libraries, and research institutes focus on theoretical exploration and upgrading their own services. Research institutions have already begun exploring the integration of knowledge organization with computational technologies to meet the needs of knowledge reuse, discovery, and value-added, achieving preliminary results in certain subfields and industries [9-11]. Comprehensive service organizations emphasize application effectiveness. For instance, the Chinese Engineering Science and Technology Knowledge Center recognized early the role of knowledge organization in big data services and vigorously promoted it [12]. Press and publishing institutions focus on reshaping information resources from the source. Driven by the former State Administration of Press, Publication, Radio, Film and Television, the National Knowledge Resource Service Center began construction, and

relevant industry standards for knowledge organization were formulated and released [13], with several publishing institutions launching specialized knowledge services based on their proprietary digital content. These research and practical efforts provide a foundation for this study.

## 1 Knowledge Organization Challenges

### 1.1 Data Expansion Challenge

The core of knowledge organization work is to standardize entities and their relationships, which may be concepts, terms, categories, etc., in different types of KOS such as thesauri, classification schemes, term systems, and ontologies. In the big data environment, the number of entities grows rapidly. For example, when the *Chinese Thesaurus* was published in 1980, it contained 91,958 formal descriptors and 17,410 informal descriptors [14], whereas the 2014 *Chinese Thesaurus (Engineering and Technology Volume)* included 196,000 preferred terms and 164,000 non-preferred terms. The 2018 *Chinese Thesaurus (Natural Science Volume)* collected 124,000 specialized terms [15]. The XLore knowledge graph constructed by Tsinghua University contains over 16.28 million entities [16].

As the scale of knowledge organization expands, relationships between different entities become increasingly rich. Traditional thesauri generally contain only relationships such as USE, UF, BT, NT, RT, and TT, with attributes including multilingual translations, definitions, scope notes, historical notes, and general notes—typically fewer than 10 types. However, the semantic network in the Unified Medical Language System (UMLS) compiled by the U.S. National Library of Medicine contains 54 types of semantic relationships [17]. The Chinese Scientific and Technical Vocabulary System developed by the Institute of Scientific and Technical Information of China features even richer semantic relationship types, with 78 secondary relationship types and 45 secondary attribute types in its new energy vehicle volume. The China Engineering Science and Technology Knowledge Center Thesaurus (Core Set) has 399 types of relationships [18], the Cyc knowledge base contains 42,500 types of relationships [19], and the XLore knowledge graph includes over 446,000 types of relationships [16].

The proliferation of entity relationship types poses challenges for KOS construction and application. In manual construction practices, knowledge engineers must confirm a relationship between two entities, and selecting from dozens of relationships takes significantly longer than choosing from a few, while accuracy and consistency decrease. In automatic construction practices, due to difficulties in handling redundant relationship types with synonymous or near-synonymous meanings, and given the large number of entities, it is nearly impossible to review each relationship for accuracy, leading to even more rapid KOS expansion. For example, in a specific corpus, analyzing the distribution of average related terms per word based on a particular relevance calculation method and different thresholds yields the results shown in Table 1 . As the threshold decreases, the average number of related terms per entry increases. For certain “star” terms

such as "teaching," the number of related terms is far greater than the average, indicating more severe expansion.

### 1.2 Document Assurance Challenge

Regardless of specific strategies, KOS construction typically considers the principle of literature assurance [20]. In practice, a document corpus is usually required as an evaluation foundation. However, the big data era faces problems of corpus incompleteness, imbalance, and inaccuracy—issues that existed previously but have become more prominent in the big data context.

**1.2.1 Incompleteness** Although computational analysis capabilities have enabled larger-scale corpora for KOS construction, the gap between corpus size and requirements has not narrowed and still fails to fully cover needs. In one practice, approximately 20 million secondary literature records were extracted as a corpus based on specific search strategies. Analysis of related terms for "immigration wave" identified "Singapore" but not "United States," a typical immigration country, which contradicts common sense. Further investigation revealed that the corpus simply did not contain sufficient co-occurrences of "immigration wave" and "United States." Additionally, large amounts of corpus data come from web crawling, while much data hidden in databases remains inaccessible, making incomplete corpora affect the underlying KOS.

**1.2.2 Imbalance** The real world is inherently unbalanced. Although corpora can undergo some screening and adjustment, they remain fundamentally unbalanced. According to Zipf's law, an empirical rule of word frequency distribution, words themselves are used unevenly in single long texts and similarly unbalanced in corpora. Due to the low frequency of certain words, large numbers of related terms obtained through statistical correlation methods become even fewer. In one practice, using the same calculation standard, "immigration wave" had 1 related term, "executive" had 43, "socialism with Chinese characteristics theory" had 298, while "teaching" had 11,128. Methods based on prior knowledge bases are similarly limited by the constraints of the knowledge providers and cannot avoid imbalance.

**1.2.3 Inaccuracy** In the internet environment, knowledge spreads faster, and errors and deviations accelerate accordingly. For example, in the scientific and technological field, the commonly used term "阈值" (threshold) is often written as "阀值" in many documents, which cannot be filtered out through statistical methods alone. In another practice, the term "executive" had a related term "ink wash painting" with a relevance score of 0.303, which defies common sense. Investigation revealed that several issues of the journal *Art Market* featured ink wash paintings by an artist named "Lin Gaoguan," but the keywords were incorrectly annotated as "executive." This error further propagated due to dissemination. Under big data resource conditions, such cases are difficult

to eliminate without manual review, and due to the volume and velocity of big data, manual review can only address a portion of cases.

### 1.3 Integration Challenge

In the big data environment, it is necessary to utilize dispersed KOS through integration and other means, but integration also introduces many problems.

**1.3.1 Inconsistency in Concept Definition** KOS focuses more on concepts but still manifests in natural language, often relying on formal matching when linking data resources. Natural language inherently has polysemy to some extent. Although terminology selection in specialized fields considers monosemy, this is difficult to guarantee in practice. During integration, linking different sources of knowledge through word forms may result in terms associated with the same word actually coming from different domains, thereby connecting originally unrelated or weakly related terms through short paths, which inevitably affects user experience in subsequent services. For example, the term "information ecology" was actually proposed separately by information science scholars and ecology scholars. Despite both being interdisciplinary fields formed by the intersection of information science and ecology, these two different connotations of "information ecology" exhibit significant differences in research objects, content, and methods [21].

Since big data resources are not all rigorous academic achievements, various abbreviations and shorthand forms are more common, especially English acronyms, making it difficult to map them to appropriate positions in KOS. For instance, "IE" is a common acronym representing different meanings in different disciplines or even within the same discipline, such as Industrial Engineering, Industrial Ecology, Ionization Energy, Information Extraction, Information Element, Information Engineering, and Internet Explorer.

According to the W3C's Simple Knowledge Organization System (SKOS) standard, mapping between different concept systems includes five mapping or alignment types: skos:closeMatch, skos:exactMatch, skos:broadMatch, skos:narrowMatch, and skos:relatedMatch [22]. These mapping types are relatively simple and may have significant semantic understanding deviations in practice, making it extremely challenging to integrate different concept systems.

**1.3.2 Inconsistency in Relationship Definition** KOS generally contains certain associative relationships. According to Chinese thesaurus construction standards, major types include equivalence relationships, hierarchical relationships (generic-specific), and associative relationships. However, relationship definitions may differ across KOS from different sources. Even in standards, several possibilities are given for relationship types: equivalence relationships include synonyms and quasi-synonyms, with synonyms having different subtypes and quasi-synonyms possibly containing antonyms and some de facto hierarchical

relationships; hierarchical relationships include genus-species, whole-part, and instance relationships; associative relationships already have 12 types listed in standards, which are not exhaustive [23-24]. Therefore, what formally belongs to a certain relationship may actually be different sub-relationships, especially in different KOS where this phenomenon is more pronounced.

**1.3.3 Inconsistency in Construction Methods** In KOS construction, there are multiple technical approaches: relying entirely on manual work, using automatic tools, or human-computer combination. Most early KOS were manually constructed. In recent years, some small-scale, rigorous KOS have relied on human-computer combination, while larger-scale KOS mainly depend on automatic tools. Automatic tools have been attempting to mimic manual work, but since the internal mechanisms of human identification and judgment cannot be fully formed into knowledge bases or comprehended by machine learning algorithms, automatic processing always produces some unexpected results. Although many automatic tool processing results can approach human levels, there will always be results that machines cannot eliminate but humans can easily identify. Therefore, different starting points, quality requirements, and construction methods naturally lead to different results, and integrating these different KOS will inevitably result in inconsistencies.

**1.3.4 Inconsistency in Knowledge Content** KOS itself avoids inconsistency and can achieve this within a small scope. However, when integrating multiple KOS into a complex, widely linked, and more broadly applicable system, fault tolerance must be considered because inconsistencies will inevitably exist and are difficult to coordinate globally. This should allow for some degree of inconsistency, such as simultaneously acknowledging that birds can fly while recognizing that a few birds like ostriches and penguins cannot. This may not degrade services but instead bring them closer to human cognition. Of course, to provide good services, non-monotonic reasoning must be implemented to match application scenarios with different knowledge content. For knowledge with fundamental errors, revisions should be made promptly based on user feedback or sampling inspections.

### 1.4 Application Challenge

**1.4.1 Diverse Requirements** Knowledge service demands are diverse. Different application scenarios require different KOS. How to integrate KOS of different scales and depths to function effectively is a difficulty in application. The ideal state might be constructing a comprehensive KOS, but this is difficult to achieve in practice due to limited human, financial, and material resources. When applying multiple KOS, attention must be paid to their coverage and depth. For example, comprehensive classification systems like the *Chinese Library Classification* cannot easily cover fine categories like “fuel cell vehicles,” and comprehensive thesauri like the *Chinese Thesaurus* cannot possibly include specialized terms like “coarse-grained steel” and describe their relationships with

other terms, let alone complex chemical compound names like "1,1,2,2,9,9,10,10-octafluoro[2.2]paracyclophane."

**1.4.2 External Adaptability** In some application practices, no suitable KOS exists, and building one from scratch is difficult. One possible solution is to borrow external KOS, but this introduces external adaptability issues. In one practice, without an appropriate thesaurus, an English financial banking domain thesaurus was translated into Chinese. During translation, some terms were difficult to translate, such as "10-K," which actually refers to "public documents (financial statements, etc., that companies must file annually with the U.S. Securities and Exchange Commission)," but without annotation, it is difficult to make a corresponding translation. Terms with American characteristics like "401K" (a type of U.S. retirement savings) may have little meaning for Chinese data organization after translation and may affect associative relationships. For example, the original relationship "cash-synonym-money" becomes "cash-synonym-cash" after translation, losing its associative value. Additionally, after translation, multiple different associative relationships may be established between two terms, mainly because the two languages cannot achieve one-to-one correspondence.

## 2 Countermeasures

In response to KOS construction and application challenges, knowledge organization work can establish a series of models from perspectives such as standardized divisible structure and asymmetric structure, multi-source updating strategies, and elastic application service models to partially address these issues.

### 2.1 Knowledge Structure

**2.1.1 Divisible Structure and Standardization** Facing big data services, KOS should not and cannot be unique; knowledge organization requires division of labor, cooperation, and sharing. In the construction of the China Engineering Science and Technology Knowledge Center, each sub-center builds its own specialized domain KOS according to business needs, while the knowledge center integrates these sub-center KOS holistically, supplements and improves them, and forms a comprehensive engineering science and technology KOS. We can assume that the ideal KOS is a large system, and a real KOS is a subsystem extracted from the large system based on divisible characteristics, which can still continue to have divisible features.

Accordingly, we propose a divisible model for KOS in big data services, as shown in Figure 1 [Figure 1: see original paper]. In terms of knowledge structure, "divisible" specifically refers to "layering, grading, blocking, and faceting." "Layering" means that the entire KOS can be divided into frequent and infrequent sets. Big data services always have some KOS data frequently accessed within a certain time period—this part is the frequent set, similar to hot data on e-

commerce or news websites. The corresponding infrequent set is similar to cold data. This layering and dynamic transformation help address user utilization of big data services and improve user experience on average. “Grading” mainly targets the frequent set, which can be further subdivided into core and extension sets that work together. The core set is relatively stable, corresponding to long-term unchanging content, while the extension set reflects timely changes. Knowledge organization content flow often occurs between the core and extension sets and between the extension and infrequent sets. In specific scenarios, KOS can have more levels, and stable values can be assigned to each piece of knowledge. “Blocking” means that KOS construction is domain-specific, with different domains equivalent to different blocks of the whole. “Faceting” means that even for a single domain, the purpose and perspective of KOS construction may differ. For the same block of new energy vehicles, some KOS may focus on policy aspects, some on technical aspects, and others on economic aspects. “Faceting” and “blocking” correspond to the professional advantages of different KOS construction and application groups, thus helping improve KOS content quality.

Divisible structures may bring potential integration risks, making standardization essential. The SKOS standard proposed by W3C is the de facto standard for KOS construction, sharing, and application on the internet, with increasingly widespread application. In the thesaurus domain, the International Organization for Standardization released the ISO 25964 standard, and China correspondingly updated the GB/T 13190 standard. In the construction of the China Engineering Science and Technology Knowledge Center and knowledge service practices in press and publishing units, relevant project or industry standards have been introduced and promoted to national standards. In classification, classification schemes such as the Dewey Decimal Classification, International Patent Classification, Chinese Library Classification, and Chinese Archives Classification have been used extensively in documents, and related construction experience can be extended to other classification schemes. Additionally, some standards have been formed, such as “SDS/T 2121-2004 Basic Principles and Methods for Data Classification and Coding,” which are helpful for KOS standardization. Of course, with the development of big data itself, the degree of standardization needs continuous strengthening.

**2.1.2 Asymmetric Structure** Thesaurus construction standards often require symmetric references for paired relationship types, and asymmetry is often treated as an important content in relationship logic verification. However, in the big data era, this approach needs adjustment. In fact, asymmetric references have long existed, such as alternate categories and formal categories in classification schemes, similar to USE/UF relationships in thesauri but not marked at formal categories [25]. Now we simply extend asymmetric references to relationship types that were previously required to be strictly symmetric.

In one practice, according to a certain standard, “teaching” had over 10,000

related terms. However, from the cognitive perspective of "teaching," it is impossible to simultaneously display more than 10,000 terms to users or include all of them in relevant programs for calculation. Among these terms, "teacher," "student," "classroom," and "multimedia" ranked high in relevance and are very important, while "mechanical professional courses," "sports game methods," and "writing knowledge" ranked lower and are related to "teaching" but not strongly associated. However, for terms like "mechanical professional courses," "sports game methods," and "writing knowledge," the relationship with "teaching" is strong. Therefore, from the perspective of building KOS, there is no need to follow the symmetry principle and establish tens of thousands of relationships for "teaching." Only the most important few or dozens of relationships need to be selected, and the same applies to other terms. This results in "mechanical professional courses," "sports game methods," and "writing knowledge" pointing to and associating with "teaching," but "teaching" does not point to these terms—instead pointing to "teacher," "student," "classroom," "multimedia," etc.—forming an asymmetric reference. This asymmetric structure is actually more common and widely accepted on social networks like Weibo, and extending it to KOS is very reasonable.

### 2.2 Multi-source Update Strategy

KOS construction is iterative. In the big data era, completely updating a KOS is difficult, so the focus is on patch updates to certain local parts of the divisible structure. The specific update model is shown in Figure 2 [Figure 2: see original paper]. Update drivers come from data resources themselves, users, and the applications built. It should be noted that these updates are often heuristic; when an update need for a certain piece of knowledge is discovered, consideration must be given to whether the update should be extended to this category of knowledge.

**2.2.1 Resource-driven Strategy** The long-term continuous change of resources is a significant feature of big data, and the pace of change is rapid. Therefore, to organize changing data resources and respond to resource changes, KOS needs to be revised and adjusted based on resources, especially the frequently changing extension set, or part of the infrequent set needs to be transformed into the frequent set. Otherwise, there will be data that cannot be organized and managed, or the level of revelation for this data will be insufficient. Whenever resources change, existing KOS need to be evaluated using coverage and other indicators. If evaluation indicators are low, the corresponding KOS needs to be updated, or it should be determined that the resource does not meet the original intention of knowledge service and needs to be deleted.

**2.2.2 User-driven Strategy** Since knowledge extracted from and used in big data services cannot be guaranteed to be 100% verified and error-free, user performance in actual use can serve as a way to discover and solve problems. Ideally, users could comprehensively feedback problems encountered in use, but

in reality, we can only analyze user behavior, especially behaviors such as not paying attention, not clicking, and relatively short attention time, to mine relevant issues for further improvement, so that users make fewer or no such errors in subsequent use. Additionally, attention should be paid to system responses to user input. If there is no response or very little response, and user input errors are excluded, the corresponding KOS likely needs adjustment. User-driven strategies actually place high demands on both user volume and user data processing capabilities. If user volume is small, reflected problems may not be representative; if data processing capabilities are weak, there will inevitably be update lags and problems that cannot be updated in time.

**2.2.3 Problem-driven Strategy** A service cannot solve all problems, so knowledge services should be relatively specialized. When the problems to be solved change, the corresponding KOS also needs to be updated. The foundation of KOS for different problems may be similar, but applications need to be customized and updated for the problem on this foundation. For example, when designing a KOS for metal materials for downstream mechanical industry researchers, more attention should be paid to mechanical property-related attributes, while for upstream metallurgical industry users, more attention should be paid to relationships between crystal structure, smelting processes, and smelting equipment. When providing services in 细分领域, gradual refinement should be based on the problem. Only in specific technical direction selection for "hybrid electric vehicles" is it necessary to distinguish between "plug-in" and "range-extended," while from a broader perspective, "hybrid electric vehicles" itself may not even need to be listed separately, instead being replaced by the higher-level "new energy vehicles."

### 2.3 Elastic Application Service Model

The mechanism of providing big data services with the help of KOS is shown in Figure 3 [Figure 3: see original paper]. Like big data resources, KOS does not directly face users but provides services to users through applications targeting different problems. Users are actually limited by both big data resources and KOS themselves. Big data resources determine the upper limit of service quantity they can obtain. Current competition in knowledge services is largely competition over data resources themselves, so various knowledge service providers tend to acquire and provide more resources, especially exclusive resources. The richness and accuracy of labeling, classification, and other knowledge organization of resources determine the upper limit of service quality. In some fields with relatively open data resources, such as patents and news, different service providers can obtain basically the same resources, so competition actually becomes competition in a series of processing and service technologies, including knowledge organization, making the elasticity of service models particularly important.

Different application services can be built on the same big data, and the cor-

responding KOS and associated resources have one or more facets reflecting different dimensions and perspectives. The specific application model of KOS in big data services is shown in Figure 4 [Figure 4: see original paper]. There are no clear boundaries between different facets. Which part of KOS a specific facet uses is dynamically adjusted according to resources, users, and the problems faced. KOS should be adaptively elastic and scalable. In extreme cases, it can include the entire KOS, but generally, the requirement for granularity of KOS is reduced to save computing and service resources.

Since KOS is iterative, there are multiple versions. When using these KOS, corresponding time labels should be used to indicate that various indexing operations were performed under specific KOS states. In fact, different parts of KOS used in an application may be from the same time or different times. For example, in patent indexing work, the International Patent Classification (IPC) is used to index invention patents and utility model patents, and the International Design Classification is used to index design patents, requiring version numbers to distinguish different classification tables used simultaneously [26]. In another archive big data project, both classification schemes and thesauri were used comprehensively for knowledge organization. The classification scheme continued to use the 1997 version of the *Chinese Archives Classification* without adjustment, while the thesaurus did not directly continue using the 1995 version of the *Chinese Archives Subject Thesaurus* but supplemented it with new subject terms reflecting archive theme changes after 1995, such as "Belt and Road" and "AIIB."

Big data services face challenges of KOS data explosion, document assurance, integration, and application. The data explosion challenge can be addressed through standardized divisible knowledge structures via multi-institution and multi-person division of labor, cooperation, and sharing. The document assurance challenge can be addressed through multi-source update strategies and the use of asymmetric structures. The integration challenge can be partially solved through standardization and divisible structures. The application challenge can be partially addressed through divisible structures combined with elastic service models and appropriate update strategies. Therefore, in the big data era, KOS itself also needs to advance with the times, forming a complex system that can be layered, graded, blocked, and faceted, with asymmetric characteristics, and real-time evolutionary iteration based on resources, users, and problems. The system may contain several relatively independent yet interrelated subsystems. In applications, appropriate subsystems need to be extracted and combined with specific data resources, problems, and users to form services, and KOS should be updated during application.

However, challenges in KOS integration and application have not been completely resolved. Real cases in other countermeasures only cover some types of KOS, and some solutions remain conceptual. These issues need to be gradually addressed in future work.

## References

[1] FIDELIA I S, GEOFFREY C B. Implications of big data for knowledge organization[J]. Knowledge Organization, 2017, 44(3): 187-198. [2] MARCIO V, MARISTELA T H, EDISON I, et al. Transforming open data to linked open data using ontologies for information organization in big data environments of the Brazilian government: the Brazilian Database Government Open Linked Data-DBgoldbr[J]. Knowledge Organization, 2018, 45(6): 443-466. [3] CHEN Chuanfu, QIAN Ou, DAI Yuzhu. Research on digital library construction in the big data era[J]. Library and Information Service, 2014, 58(7): 40-45. [4] WANG Yuefen, FU Zhu. Application of knowledge representation and organization methods in big data environments[J]. Digital Library Forum, 2014(3): 32-43. [5] YANG Xiaolan, QIAN Cheng, ZHU Fuxi. A cloud computing-based evaluation method for big data service resources[J]. Computer Science, 2018, 45(5): 295-299. [6] JIANG Yongfu. On knowledge organization[J]. Library and Information Service, 2000, 44(6): 5-10. [7] HE Defang, QIAO Xiaodong, ZHU Lijun, et al. Chinese scientific and technical vocabulary system (new energy vehicle volume)[M]. Beijing: Scientific and Technical Documentation Press, 2012. [8] LI Xuhui, FAN Meihui. Knowledge association in big data[J]. Information Studies: Theory & Application, 2019, 42(2): 68-73, 107. [9] LI Xuhui, QIN Shuqian, WU Yanqiu, et al. Knowledge organization in large-scale data from a computational perspective[J]. Document, Information & Knowledge, 2018(6): 94-102. [10] SUN Tan, LIU Zheng, CUI Yunpeng, et al. Exploring a new generation of open knowledge service architecture integrating knowledge organization and cognitive computing[J]. Journal of Library Science in China, 2019, 45(3): 38-53. [11] LU Quan, JIANG Chao, CHEN Jing. Research on electronic medical record big data organization based on extended disease ontology[J]. Document, Information & Knowledge, 2019(1): 109-118. [12] PAN Gang, ZHANG Yunliang, ZHONG Qinghong. Thoughts and practices on knowledge services in engineering science and technology[J]. Technology Intelligence Engineering, 2018, 4(5): 4-12. [13] Digital Publishing Department, State Administration of Press, Publication, Radio, Film and Television. Notice on approving and releasing 8 project standards including "Knowledge Service Standard System Table"[EB/OL]. [2019-04-30]. http://www.gapp.gov.cn/ztzzd/zdgzl/cbyszhzsjxmxzl/content/4384/274644.shtml. [14] HE Defang. Review and prospect of the Chinese Thesaurus[J]. Information Studies: Theory & Application, 2010, 33(2): 1-4. [15] Chinese Thesaurus (Natural Science Volume)[M]. Beijing: Scientific and Technical Documentation Press, 2018. [16] TSINGHUA UNIVERSITY KEG. XLore[EB/OL]. [2019-04-30]. https://xlore.org/. [17] NATIONAL LIBRARY OF MEDICINE (US). UMLS reference manual[EB/OL]. [2019-04-30]. https://www.ncbi.nlm.nih.gov/books/NBK9679/. [18] China Engineering Science and Technology Knowledge Center. China Engineering Knowledge Center thesaurus core set[EB/OL]. [2019-08-09]. http://data.ckcest.cn/CKT-C. [19] DOUGLAS L. Cyc[EB/OL]. [2019-04-30]. https://en.wikipedia.org/wiki/Cyc. [20] LIU Han. On the principle of literature assurance in information organi-

zation[J]. Journal of the National Library of China, 2019, 28(1): 57-65. [21] LOU Cequn, GUI Xiaomiao, YANG Xiaoxi. Thoughts on the construction of information ecology discipline in China[J]. Information Science, 2013, 31(2): 13-18. [22] W3C WORKING GROUP. SKOS simple knowledge organization system primer[EB/OL]. [2019-04-30]. https://www.w3.org/TR/skos-primer/. [23] National Documentation Standardization Technical Committee. Guidelines for the establishment and development of Chinese thesauri: GB 13190-91[S]. Beijing: Standards Press of China, 1992. [24] National Documentation Standardization Technical Committee. Information and documentation—Thesauri and interoperability with other vocabularies—Part 1: Thesauri for information retrieval: GB/T 13190.1-2015[S]. Beijing: Standards Press of China, 2015. [25] WANG Xiaohua. Review of alternate categories in the 5th edition of the Chinese Library Classification[J]. Journal of Library Science, 2017, 37(11): 132-134. [26] Patent Office, China National Intellectual Property Administration. China patent literature cataloging standards[EB/OL]. [2019-04-30]. http://www.sipo.gov.cn/wxfw/zlwxxxxggfw/zsyd/bzyfl/zlwxyxxbz_{gnbz}/1053740.htm.

*Note: Figure translations are in progress. See original paper for figures.*