Evolution of Research on Vocabulary Semantic Organization (1998-2018) Postprint

Authors: Tao Jun

Date: 2023-08-27T00:00:00+00:00

Abstract

[Purpose/Significance] The semantic organization of vocabularies constitutes an important component of research on the semanticization of library collections. Reviewing the historical evolution of this field is conducive to clarifying key foci and promoting its better development. [Method/Process] Based on differentiating and analyzing the core terminology in the field of vocabulary semantic organization, this study proposes a research framework of "standards and specifications—semantic organization methods—supporting technologies—vocabulary applications," and reviews representative literature on Chinese thesaurus semantic research based on this framework. [Results/Conclusion] This paper puts forward the definition and main framework of vocabulary semantic organization, reveals the core concepts of vocabularies, ontologies, linked data, and their organic connections; takes thesauri as an example to review representative research work on vocabulary semantic organization in China over the past decade; compares the intrinsic similarities and differences between traditional vocabulary research and semantic organization research; and provides a review and outlook for vocabulary semantic organization research in China.

Full Text

Preamble

Vol. 62 No. 21 November 2018
ChinaXiv Cooperative Journal

The Evolution of Research on Semantic Organization of Vocabularies (1998-2018)

School of Public Management, Northwest University, Xi'an 710127

Abstract

[**Purpose/Significance**] Semantic organization of vocabularies constitutes a crucial component of collection semantic research. Reviewing the historical evolution of this field helps clarify key priorities and promotes its better development. [**Method/Process**] Based on the analysis of core terminology in the field of vocabulary semantic organization, this paper proposes a research framework of "standards and specifications—semantic organization methods—supporting technologies—vocabulary applications." Using this framework, the paper reviews representative Chinese literature on thesaurus semantic research. [**Result/Conclusion**] The paper puts forward a definition of vocabulary semantic organization and its main framework, revealing core concepts such as vocabularies, ontologies, and linked data and their organic connections. Taking the thesaurus as an example, it summarizes representative research work on Chinese vocabulary semantic organization over the past decade. It compares the intrinsic similarities and differences between traditional vocabulary research and semantic organization research, and provides commentary and prospects for Chinese vocabulary semantic organization research.

Keywords: vocabulary, linked data, semantic web, resource description framework
Classification Number: G254.2
DOI: 10.13266/j.issn.0252-3116.2018.21.017

The Linked Open Data (LOD) cloud demonstrates that RDF datasets covering literature, biology, geography, and other domains are proliferating rapidly, and the World Wide Web is evolving toward a data web containing massive conceptual entities and rich semantic relationships between them. The advancement of semantic search projects such as Google Knowledge Graph and Baidu's "Zhixin" has shifted semantic web research and practice from a solo performance in traditional academic fields to a collaborative symphony resonating between academia and industry [**?**]. Alongside the development of the semantic web, knowledge organization is undergoing a transformation from traditional tool-assisted positioning to data-intelligent research. Particularly since the launch of the linked data movement, the semantic research of various library resources has garnered widespread attention within the profession. As a fundamental tool for collection resource indexing and assisted retrieval, vocabulary semantic organization research constitutes an essential component of collection semantic research.

There is not yet a unified definition of vocabulary semantic organization. This paper defines it as the application of semantic web-related standards and web engineering technologies to advance the description, linking, and application of vocabularies in network environments. As an interdisciplinary field, previous research includes several reviews: Song Wen et al. reviewed research on vocabulary mapping, interoperability, and conversion to ontologies [**?**]; Xue Chunxiang et al. examined legacy resource linked data publishing [**?**], terminology services, and RDF linking across different datasets [**?**]. The most recent and closely

related work is M.L. Zeng et al.'s "Knowledge organization system in the semantic web: a multi-dimensional review," which explores the role of vocabularies in linked open datasets from multiple dimensions including dataset producers, vocabulary producers, and vocabulary users [**?**]. However, these studies either focus on narrow domains lacking close thematic connection to vocabulary semantic organization, or they are not structured around the logical framework of vocabulary semantic organization evolution. Moreover, because vocabulary semantic organization spans both semantic web and vocabulary themes, traditional single-theme reviews tend to fragment the inherent connections between cross-cutting topics, making it difficult to understand the position and role of related work within the overall development pattern. Therefore, this paper attempts to analyze the evolution of key concepts and their intrinsic similarities and differences from a more open perspective, while relying on the vocabulary semantic organization framework to selectively review representative research work in China's vocabulary semantic organization field.

The contributions of this paper are mainly threefold: It proposes a definition and main framework for vocabulary semantic organization, revealing core concepts such as vocabularies, ontologies, and linked data and their organic connections; It summarizes representative research work on Chinese vocabulary semantic organization over the past decade, using the thesaurus as an example; It compares the intrinsic similarities and differences between traditional vocabulary research and semantic organization research, and provides commentary and prospects for Chinese vocabulary semantic organization research, particularly summarizing term mapping methods based on word form, structure, and corpus in vocabulary interoperability.

2. Concept Analysis

Over the past two decades of rapid development, vocabulary semantic organization research has not only expanded the conceptual and categorical systems related to vocabularies but has also seen iterative updates in technical standards and methods supporting vocabulary semanticization, such as semantic web, web engineering, and artificial intelligence technologies. This conceptual expansion and evolution of supporting technologies have enhanced the field's professionalism but also risk confusion in terminology and conceptual relationships, potentially slowing domain research. Therefore, clarifying relevant concepts and their internal logic is crucial for promoting field development.

2.1 Vocabulary

Vocabulary has both narrow and broad definitions. In the narrow sense, vocabulary refers to controlled vocabularies, also known as thesauri, which in China typically means subject heading lists [**?**]. With the development of network environments, the connotation and extension of vocabulary concepts continue to expand. In the broad sense, vocabulary includes authority files, classification schemes, thesauri, semantic networks, and ontologies (see Figure 1). American

digital library expert G. Hodge termed these knowledge organization systems (KOS) in 2000 [**?**], marking the transition of traditional dispersed literature organization tools toward intensive development in network environments.

As mentioned, the vocabulary concept developed from the term "controlled vocabulary." Controlled vocabularies contain a series of terms and display different relationship types, with the essential characteristic being the expression of terms/concepts and their inter-term relationships. In 2005, the ANSI/NISO Z39.19-2005 standard released by the American National Standards Institute expanded the definition of controlled vocabularies, categorizing them by control level from weak to strong into pick lists, synonym rings, taxonomies, and thesauri. In 2011, the W3C Library Linked Data Incubator Group divided library linked datasets into RDF element sets and value vocabularies [**?**], also called structured vocabularies and concept vocabularies [**?**]. The former typically includes MARC metadata, DC, RDA, and Bibframe, while the latter refers to broad vocabularies centered on thesauri containing hierarchical, associative, and equivalence relationships. Vocabulary extension has expanded significantly from original normative terms and relationships to expressing domain concepts and their relationships. Major vocabularies and their semantic representation standards are shown in Table 1 .

Since computers were applied to documentation and information work, vocabulary description formats have evolved from electronic to semantic representation. Electronic formats included MARC in database environments and HTML for web pages. For example, the AGROVOC thesaurus transitioned from print to electronic storage in relational databases in 2000 , and China implemented MARC representation for the Chinese Classification Thesaurus in 2005. However, HTML merely digitized traditional text and could not support computer semantic representation and processing. While MARC could reveal semantic information through metadata, its standards could not meet network environment requirements for data openness, sharing, and web processing. In 1998, XML became the standard format for data representation in web environments, achieving separation between semantic information (metadata) and data content while meeting network interaction and sharing needs. Consequently, the Library of Congress and other institutions promoted the transition from MARC to MARCXML and MARC21, with various metadata standards and vocabularies gradually adopting XML as their preferred language. To support intelligent computer processing of online resources, electronic vocabularies continued evolving toward intelligent semantic representation.

2.2 Semantics

Vocabulary semanticization revolves around semantic web standards. This paper defines semantics as comprising conceptual relationship logic, formal representation based on this logic, and intelligent mechanisms. The semantic web's core is achieving semantics through ontologies. Ontology contains two meanings: First, the ontology model. An ontology is a model of concepts and their

relationships abstracted from a domain, typically describing systematic concepts and relationships [**?**]. Properties of ontology model concepts are called data properties, while properties of inter-concept relationships are called object properties. Atomic concept statements in ontology models consist of RDF triples of "subject-predicate-object." Second, ontology formalization. Transitioning from ontology models to computer processing requires support from a series of web language specifications. XML, RDF, and OWL are semantic representation specifications for web data launched by the W3C (see Table 1). Based on these languages, multiple different semantic descriptions can be constructed for ontology model concepts and properties.

SKOS is a representation specification specifically proposed for the relatively simplified structure of vocabularies, distinguishing it from ontologies. China implemented SKOS representation for the Chinese Classification Thesaurus in 2010 [**?**]. Linked data, proposed by Tim Berners-Lee in 2006, is a technical standard [**?**] whose goal can be summarized as achieving computer modeling, representation, and association discovery of various concepts/entities and their relationships. Linked data must follow two foundational standards: First, concepts in linked data must be represented using HTTP URIs, enabling each concept to be accessed via HTTP protocol for open data sharing in web environments; second, providing rich URIs to discover or link more concepts/entities. As more individual linked datasets are published, linked data's greater significance lies in building associations between identical or related concepts/entities across different linked datasets, with relationship vocabularies like owl:sameAs and rdfs:seeAlso supporting concept entity linking.

2.3 Relationships Among RDF, Ontology, and Linked Data

RDF is the modeling standard for semantic web data representation; both ontology and linked data records must follow RDF models to achieve structuring. Generating linked data from a data resource requires establishing an ontology model for the resource system and formalizing the ontology model; linked data is essentially instances of ontology models. The differences between ontology and linked data mainly manifest in the transformation of semantic construction mechanisms across two dimensions: First, different semantic development philosophies. Early semantic web development focused on independently constructing ontologies and achieving richer semantics through more powerful reasoning logic. After linked data's proposal, the emphasis shifted to reusing structured resources including vocabularies to build lightweight ontologies [**?**], while relying on entity associations across different resources to enrich semantics, downplaying original ontology construction and complex reasoning logic. Second, different ontology model formalization philosophies. Unlike traditional ontology models that focused on independently defining a single ontology vocabulary set for formalization, linked data emphasizes prioritizing the use of multiple mature ontology vocabulary sets to formalize ontology models. Maximizing the reuse of mature ontology vocabulary sets helps lay the foundation for

forming association relationships with other RDF datasets later. Due to differences in ontology vocabulary set selection, ontology models established around a data resource may have multiple different ontology formalization approaches.

3. Research Content

Based on the definition of vocabulary semantic organization, this paper divides the vocabulary semantic organization framework into four layers from standards and specifications to vocabulary application (see Figure 2). The basic logic is that vocabulary semantic organization is the framework's core, and its implementation methods and concepts have deepened dynamically alongside evolving standards over the past two decades.

Supporting technologies in Figure 2 include but are not limited to the above modules for two reasons: First, technology always serves content, and vocabulary semantic organization content continues enriching alongside technological development; second, due to different scholars' technical classifications and varying content emphases, supporting technologies may take diverse forms. The following sections review Chinese thesaurus semantic organization work using the Figure 2 framework, focusing on reflecting main work while revealing organic connections across different domains.

3.1 Vocabulary Semantic Organization

Vocabulary semantic organization developed progressively based on vocabulary databasing and web page representation, oriented toward semantic web-related standards [**?**]. Following Figure 2, the author divides vocabulary semantic organization research into three evolutionary layers: "semantic description—ontology conversion—linked data publication" (see Table 2). Classification schemes, authority files, and thesauri all include these processes. Comparatively, thesauri occupy the relationship layer within the overall vocabulary structure, making them more representative. The following sections focus on reviewing representative work on thesaurus semantic organization, with research on Chinese classification schemes and authority files covered in literature [**?**].

First, from an early development perspective, since semantic-related standards were not yet mature, the library and information science community focused on understanding semantic-related standards and concepts through thesauri [**?**]. Thesauri formed in the print era, centered on terms with coarse-grained relationships like "use, replace, broader, narrower, related," bear some similarity to ontologies formed in network environments that are concept-centered and emphasize fine-grained semantic relationships. Consequently, using thesauri to build domain ontologies or enable semantic description became central research topics.

Second, from a research object perspective, the vocabulary semantic description stage focused on theoretical exploration of local term units due to immature semantic standards and technologies. In the linked data stage, with mature

SKOS standards and linked data publication technologies, research shifted to holistic vocabulary exploration based on database or web versions. The Chinese Classification Thesaurus and Agricultural Science Thesaurus, being ahead in electronic construction, laid foundations for linked data practice.

Third, from a semantic organization hierarchy perspective, the three layers have progressive relationships. Semantic description focuses on vocabulary semantic representation without structural adjustment. Converting vocabularies to ontologies involves not only semantic representation but also redefining conceptual models according to different contextual needs. Publishing vocabularies as linked data incorporates previous semantic description, ontology modeling, and formalization processes, but differs by integrating linked data standards and treating vocabularies as entity instances alongside other bibliographic resources. This reflects how vocabulary functions are evolving beyond traditional tool-assisted positioning toward data intelligence under semantic technology applications.

Comprehensive analysis shows vocabulary semantic organization is a continuously dynamic development process. Based on this foundation, the following sections review key research on Chinese thesaurus semantic conversion. Scholars including Zeng Xinhong, Liu Libin, Duan Rongting, Xian Guojian, Liu Huamei, and Ou Shiyan have conducted methodological explorations of Chinese thesaurus semantic conversion, representing influential work in the field (see Table 3).

From the perspective of semantic representation languages and historical development stages, before the SKOS standard specifically for vocabulary conversion emerged, Zeng Xinhong explored semantic description of large-scale general subject heading lists using OWL language [**?**]. Research focused on major subject heading lists like the Chinese Classification Thesaurus, Chinese Archives Subject Thesaurus, and Agricultural Science Thesaurus, which had been 电子化 (electrified), providing solid foundations for semantic exploration. After SKOS became a recommended vocabulary construction standard in 2009, subsequent research primarily used SKOS.

From a conversion content perspective, Zeng Xinhong, Liu Libin, and Duan Rongting pioneered thesaurus conversion research. Zeng Xinhong and Duan Rongting provided overall conversion schemes combining OWL and SKOS languages for the Chinese Classification Thesaurus and Chinese Archives Subject Thesaurus, including SKOS description of main tables, auxiliary tables, and indexes [**?**, **?**]. Liu Libin et al. focused on automatic conversion exploration using core relationships like “use, replace, broader, narrower, related” [**?**]. Xian Guojian et al. concentrated on the Agricultural Science Thesaurus, building an associated data publication platform simultaneously [**?**].

From semantic relationship and technical implementation perspectives, Liu Libin, Xian Guojian, and Ou Shiyan used the Java programming language to achieve automatic SKOS conversion of Chinese thesauri. Liu Libin et al. were

the earliest to conduct Chinese thesaurus semantic conversion, implementing automatic conversion of core relationships including "use, replace, broader, narrower, related, and top term" for the Chinese Classification Thesaurus [**?**]. Xian Guojian et al. used SKOS and SKOS-XL for semantic representation of the Agricultural Science Thesaurus and implemented linked data publication based on Virtuoso [**?**]. Ou Shiyan, building on traditional relationship SKOS conversion, used SKOS-EX to represent complex concepts like coordination, facets, and top terms, implemented full description and batch conversion of the Chinese Classification Thesaurus vocabulary portion through Java, and published linked data based on the Pubby platform [**?**]. Liu Huamei proposed a mapping scheme converting Chinese Classification Thesaurus MARC data to SKOS and implemented batch conversion of subject concepts using VB language [**?**].

3.2 Supporting Technology Research

Publishing vocabularies as linked data primarily relies on web engineering technologies, focusing on linked data publication technology, visualization technology, and linking matching technology.

3.2.1 Linked Data Publication As a type of RDF dataset, various RDF dataset linking publication methods apply to vocabularies [**?**]. Generating thesaurus RDF datasets requires consideration of vocabulary construction foundations and structural differences. Steps can be summarized as: HTTP URI determination—ontology modeling based on vocabulary structure—entity RDFization—entity linking—RDF files—linked data publication—open SPARQL query. From a relational database file input-output perspective, outputs include text files—SQL files—RDF files (including rdf/xml, owl, skos files, etc.), as shown in Figure 3 [Figure 3: see original paper].

Taking Chinese thesaurus conversion as an example, Xian Guojian et al. [**?**] used the Agricultural Science Thesaurus relational database as a foundation, forming relevant schemes based on its structure, including setting HTTP URIs and mapping main relationships like "use, replace, broader, narrower, related" to SKOS tags. Liu Libin and Ou Shiyan et al. [**?**, **?**] used the Chinese Classification Thesaurus web version as a foundation, crawling HTML formats to obtain large term sets, preprocessing to obtain text files, writing the thesaurus text files into SQL databases, determining ontology modeling schemes, and using toolkits like Jena API to write Java conversion programs for batch conversion to RDF datasets.

3.2.2 Visualization RDF datasets excel at reflecting fine-grained concepts and multiple semantic relationships, but their serialization formats focus on machine processing, making it difficult for humans to effectively identify potential relationships. Thus, visualization is essential. Open-source ontology visualization tools like WebVOWL, Protégé, and Welkin help reveal semantic relationships between vocabulary concepts. For example, Fan Wei et al. used Graphviz

and Protovis libraries to visualize linked data in a prototype terminology service system built from Chinese Classification Thesaurus subject term data [?]. Hong Na et al. compared five visualization tools—RelFinder, Graphviz, RDFGravity, RDFViz++, and Gruff—from perspectives including development platforms, application types, open-source status, input-output formats, triplestore support, and interaction capabilities [?]. Lower-threshold tools are more popular; Hong Na, Ren Ruijuan, and Shi Zeshun et al. used RelFinder to explore terminology service applications in biomedicine, CNKI, and LISA databases, explaining application scenarios like data creation, information retrieval, and resource navigation [?]. Zhao Longwen and Chen Tao used Gruff to study government domain linked data visualization and genealogy linked data RDF visualization [?].

3.2.3 Linking Matching Building links between different RDF datasets is a requirement of the linked data five-star standard. As of February 2017, only 10% of data in the LOD cloud diagram had achieved five-star linked data status [?]. Following the 2006 mapping between China's Agricultural Science Thesaurus and AGROVOC [?], the UN Food and Agriculture Organization explored linking AGROVOC with EUROVOC, NALT, GEMET, STW, LCSH, and RAMEAU and publishing them as linked data in 2011 [?]. Compared to overseas research, Chinese exploration in this area is minimal, with almost no research on linking between large-scale thesauri. Tao Jun [?] and Zhu Wenjing et al. [?, ?] introduced automated tools for discovering links between different RDF datasets, including SILK. Xian Guojian et al. briefly introduced precise matching results between the Agricultural Science Thesaurus and AGROVOC, NALT, LCSH, and EUROVOC when exploring linked data publication, but did not elaborate on experimental processes using relevant mapping tools [?]. More research focuses on linking bibliographic datasets or literature resource datasets [?]. For example, Yu Weili et al. used overseas bibliographic data fragments integrated with SILK to find equivalence relationships, exploring links between bibliographic datasets and DBpedia [?]. Zhong Yuanxin and Liu Wei et al. discussed linking author information between Shanghai Library bibliographic data and DBpedia [?].

3.3 Vocabulary Semantic Application Research

Chinese exploratory research on vocabulary semantic applications mainly manifests in terminology services and semantic knowledge bases. From a library and information science perspective, vocabulary construction aims to support terminology services in network environments, primarily including web interfaces for human-accessible term queries and application programming interfaces for computer processing. From a semantic web construction perspective, publishing vocabularies as linked data essentially creates a refined semantic knowledge base whose conceptual relationships can support knowledge discovery. From an application domain perspective, terminology services and knowledge bases can be applied across multiple fields including medicine, biology, and law.

3.3.1 Terminology Services Multiple scholars have explored terminology service technical implementation based on REST architecture. Ou Shiyan et al. implemented a REST-based terminology service prototype system using the Chinese Thesaurus as an example, explaining terminology service application forms in cataloging, metadata creation, information retrieval, and resource navigation [**?**]. Zeng Xinhong et al. provided terminology services based on web service APIs and web page retrieval using the classification system CLSS [**?**]. Additionally, some scholars built terminology service prototype systems based on SKOS or OWL documents. Xu Lei and Dong Hui used the NCI Cancer Thesaurus OWL document from the U.S. National Cancer Institute as a data source, building a REST-based terminology service using the graph database Neo4j as a storage platform [**?**]. Fan Wei et al. built a terminology service prototype based on a linked data publication framework using Cherrypy + TDB + Joseki [**?**].

3.3.2 Knowledge Bases Some scholars have explored the role of vocabularies in supporting conceptual retrieval around semantic knowledge base construction. Wang Jun from Peking University built a bibliographic ontology model KVision and formalized it based on Chinese Classification Thesaurus categories, subject terms, and metadata, using Peking University Library's computer science bibliographic data as ontology instances to construct a semantic knowledge base enabling conceptual retrieval [**?**]. Ou Shiyan et al. implemented RDF linking of controlled vocabularies, personal names, and place names from multiple literature data sources through linked data conversion [**?**], and further explored using natural language processing to transform natural language into structured SPARQL queries for integrated searching and automatic question-answering across multiple RDF datasets [**?**].

4. Discussion

The above sections reviewed representative research in Chinese vocabulary semantic organization across four aspects, explaining each module's function and progressive relationships between modules. Further discussion follows from scientific questions and research characteristics.

4.1 Internal Connections in Vocabulary Semantic Organization Research

In traditional library and information science, three typical vocabulary research problems are vocabulary standards and specifications, vocabulary construction and updating, and vocabulary term mapping (see Table 4). Under semantic web standards, the focus shifts to exploring issues like "vocabulary semantic description—vocabulary conversion to ontology/linked data publication—open linking between different datasets." These new questions build upon the aforementioned classic problems. First, from a terminology dimension, semantic representation of concepts and attributes depends on applying various seman-

tic specifications while requiring deep understanding of vocabulary standards. Second, from an output dimension, converting vocabularies to ontologies and publishing them as linked data essentially promotes computer processing, which requires rich vocabularies and term relationships as foundations—the goal of vocabulary construction and updating [**?**]. Third, from a relationship dimension, open linking between different datasets focuses on exploring equivalence relationships with other RDF datasets, consistent with vocabulary mapping's exploration of term or concept mapping. Although technical implementation may use linking discovery methods for linked datasets like Silk, underlying mapping or alignment methods still primarily rely on various string similarity algorithms and APIs [**?**], consistent with ontology matching methods. In summary, vocabulary semantic organization and vocabulary construction are interconnected wholes.

4.2 Strong Practical Orientation but Limited Specialized Research Teams

Overall, Chinese vocabulary research demonstrates two main characteristics:

(1) Strong practical orientation. Vocabulary research primarily involves institutions supported by the National Science and Technology Library, emphasizing engineering practice. For example, the National Science Library of Chinese Academy of Sciences has conducted extensive work on foreign knowledge organization platform and integrated system construction [**?**]; the Institute of Scientific and Technical Information of China has advanced national thesaurus database construction based on Chinese Thesaurus updates, forming strong practical characteristics [**?**]; the Chinese Academy of Medical Sciences has long tracked and explored semantic applications in medical vocabularies; the Chinese Academy of Agricultural Sciences has developed distinctive features in agricultural science thesaurus linked publication and platforms [**?**]. The National Library and multiple library and information institutions nationwide have formed influence around open research on the Chinese Classification Thesaurus. However, most institutions conduct free exploration without aggregation, resulting in comprehensive practice tracking and overview studies being predominant while engineering research and practical innovation around specific scientific problems are relatively lacking, showing popularized and fragmented characteristics.

(2) Limited specialized research teams. Compared to other research areas, continuous research scale by library practice departments and LIS scholars on vocabularies is relatively insufficient. This phenomenon has internal and external factors. Internally, the professional threshold for vocabulary semantic organization research is gradually rising. As vocabulary standards adapt to network environments, vocabulary research has generated numerous new concepts and terms, while semantic web, web engineering, and natural language processing technologies increasingly dominate vocabulary development. Interdisciplinary integration makes vocabulary research no longer belong to traditional LIS knowledge domains but increasingly depend on computer application sup-

port. Externally, as a core LIS domain, the traditional inherent positioning of vocabularies is gradually marginalizing in network environments. Both factors are reducing continuous vocabulary researchers, and the vocabulary field is undergoing potential changes with AI technology application and semantic search service formation, bringing both challenges and opportunities for vocabulary semantic organization research.

5. Research Outlook

Vocabulary semantic organization developed gradually to meet network era needs. As vocabularies form RDF datasets and are published as linked data, their applications extend beyond traditional literature retrieval and assisted semantic indexing terminology services. As search engines potentially transform toward intelligent retrieval, they remain foundational tools for network survival. Therefore, future research should optimize terminology service practice while exploring vocabulary applications around semantic search as a deepening focus. Specifically, three aspects warrant attention:

(1) Strengthen large-scale vocabulary semantic linking exploration. With semantic environment formation, one direction for semantic search development is toward information association. Conducting 对照映射 (comparison mapping) between large thesauri like the Chinese Thesaurus and Chinese Classification Thesaurus with domestic and foreign counterparts [**?**, **?**], and promoting faceted transformation of the Chinese Classification Thesaurus will support deep retrieval based on information discovery, enhancing large thesaurus application value in the new era. On this basis, further exploring linking methods and practices between different vocabulary types, especially open linking experiments and technical innovation research between vocabularies and various RDF datasets (vocabulary datasets and other resource datasets), represents a future priority [**?**].

(2) Broaden technical scope of vocabulary semantic organization [?]. Current LIS vocabulary semantic organization technologies concentrate on W3C-advocated linked data technologies. From a development trend perspective, linked data standards and technologies are just one branch. Correspondingly, rapid development and multi-domain application of artificial neural network models and cognitive computing have made these methods powerful tools for developing data intelligence. Obviously, semantic web methods based on manually establishing knowledge representation models through ontologies and metadata are being challenged by intelligent computing methods achieving full-process unsupervised intervention based on word and sentence vectors. As Elsevier's Chief Architect B.P. Allen noted, the semantic web is human-based rather than machine-based, with shortcomings in helping machines learn to read, requiring future knowledge graph construction through machine reading [**?**]. Therefore, exploring knowledge graph methods based on relevant platform systems or API applications, absorbing database, natural language processing, and machine learning (deep learning) multidisciplinary fields, represents an

important means for deepening vocabulary semantic organization research [**?**].

(3) Deepen application domains of vocabulary semantic organization. First, further promoting linked data publication of various resources is fundamental to deepening semantic search. By building RDF links between vocabularies and domain datasets, vocabulary concepts become aggregation intermediaries for various domain datasets. In other words, the more domain datasets are published with linked data, the broader the potential aggregation surface for vocabulary datasets. The UK, Finland, and others have further promoted linked publication of historical, legal, and other resources through digital humanities movements, exemplifying this work. Second, combining metadata vocabularies to explore broader application scenarios in commerce, social networks, and intelligent transportation for people, geography, and applications [**?**], and leveraging vocabulary semantic annotation to study user profiling and personalized recommendation in big data environments are important aspects for expanding vocabulary applications.

Since the establishment of the NKOS (Network Knowledge Organization System) group in 1998, vocabularies have continuously evolved from networkization to semanticization, with vocabulary semantic organization research achieving substantial development in concepts, content, methods, and technologies. Compared to traditional review studies focusing on single domains like vocabularies or semantic webs, this paper combines vocabularies with semantic web standards and outlines the overall development and internal mechanisms of vocabulary semantic organization longitudinally, compensating for shortcomings in historical similar research. However, this paper also has limitations. First, its large content span results in slightly insufficient local horizontal analysis while highlighting key points, such as limited introduction to vocabulary standards research and domain applications of vocabulary semanticization. Second, regarding vocabulary semantic organization research work, this paper mainly focuses on Chinese vocabulary semanticization work, while overseas similar research is equally important for deepening vocabulary studies. These shortcomings will be addressed in future research.

References

[1] Wang Haofen. Large-scale knowledge graph technology[J]. Communications of CCF, 2014, 10(3): 64-68.
[2] Zou Lei. Research and development trends of knowledge graphs[J]. Communications of CCF, 2017, 13(8): 49-54.
[3] Chen Chen, Song Wen. Review of thesaurus mapping research[J]. Library and Information Service, 2012, 56(12): 113-119.
[4] Duan Ruilong, Song Wen. Review of methods for converting thesauri to ontologies[J]. Journal of Intelligence, 2012, 31(7): 66-71.
[5] Song Wen. Research on semantic interoperability of knowledge organization systems[J]. Library Tribune, 2012(11): 117-121.
[6] Xue Chunxiang, Qiao Xiaodong, Zhu Lijun. Review of term mapping in

KOS interoperability[J]. New Technology of Library and Information Service, 2010, 26(2): 31-37.
[7] Marjit U, Sharma K, Sarkar A, et al. Publishing legacy data as linked data: a state of the art survey[J]. Library Hi Tech, 2013, 31(3): 520-535.
[8] Ou Shiyan. Review of foreign terminology registries and terminology services[J]. Journal of Library Science in China, 2014, 40(5): 110-126.
[9] Tao Jun. Comprehensive analysis of RDF linking framework based on Linked Data[J]. New Technology of Library and Information Service, 2011(12): 1-8.
[10] Golub K, Tudhope D, Zeng M L, et al. Terminology registries for knowledge organization systems: functionality, use, and attributes[J]. Journal of the Association for Information Science and Technology, 2014, 65(9): 1901-1916.
[11] Zeng M L, Philipp M. Knowledge organization system (KOS) in the semantic Web: a multi-dimensional review[EB/OL]. [2018-05-05]. https://arxiv.org/abs/1801.04479.
[12] Dai Weimin. Information Organization[M]. 2nd ed. Beijing: Higher Education Press, 2009: 114.
[13] Hodge G. Systems of knowledge organization for digital libraries: beyond traditional authority files[R/OL]. [2017-12-20]. https://www.clir.org/wp-content/uploads/sites/6/pub91.pdf.
[14] Zeng M L, Fan W. SKOS and its application in transferring traditional thesauri into networked knowledge organization systems[EB/OL]. [2017-10-18]. https://www.researchgate.net/publication/265817740_{{{SKOS}}*{{and}}*{{{Its}}*{{Application}}*{{{in}}}
[15] W3C. Library data resources[EB/OL]. [2017-10-15]. https://www.w3.org/2001/sw/wiki/LLD/Library_{{{
2011.
[16] NISO. Vocabulary management[EB/OL]. [2018-02-20]. http://www.niso.org/standards-committees/vocab-mgmt.
[17] Bawden D, Robinson L. Introduction to Information Science[M]. London: Facet Publishing, 2012: 1-22.
[18] Fan Wei. Toward open, semantic, and linked Chinese Classification Thesaurus[J]. Journal of Library and Information Science (Taiwan), 2017, 43(1): 155-170.
[19] Bizer C, Heath T, Berners-Lee T. Linked Data-the story so far[J]. International journal of semantic Web information system, 2009, 5(3): 1-22.
[20] Ji Shanshan, Liu Zheng, Song Wen. Key technologies for thesaurus reconstruction to ontology[J]. Library and Information Service, 2017, 61(2): 103-108.
[21] Wang Jun, Zhang Li. Research status and development trends of network knowledge organization systems[J]. Journal of Library Science in China, 2008, 34(1): 65-69.
[22] Zhang Shinan, Song Wen. Design of SKOS description scheme for "Ke Tu Fa"[J]. New Technology of Library and Information Service, 2010(6): 7-11.
[23] Zhao Jie, Jia Junzhi. Construction and development of Chinese name authority files in data networks[J]. Library and Information Service, 2013(1): 8-12.
[24] Jia Junzhi. Considerations on converting Chinese Thesaurus to ontology[J].

Journal of Library Science in China, 2007, 33(4): 41-44.
[25] Mao Jun. Research on thesaurus based on RDF[J]. Journal of the China Society for Scientific and Technical Information, 2003, 22(2): 163-168.
[26] Zeng Xinhong. OWL representation of Chinese Classification Thesaurus and its deep semantic revelation[J]. Journal of the China Society for Scientific and Technical Information, 2005, 24(2): 151-160.
[27] Liu Libin, Zhang Shouhua, Pu Demin, et al. Automatic SKOS description conversion of Chinese Classification Thesaurus[J]. Journal of Library Science in China, 2009, 35(6): 56-60.
[28] Duan Rongting. Research on semantic network application of Chinese Archives Subject Thesaurus[J]. Archives Science Study, 2010(6): 66-70.
[29] Duan Rongting. Research on subject thesaurus semantic networking based on Simple Knowledge Organization System: taking Chinese Archives Subject Thesaurus as an example[J]. Journal of Library Science in China, 2011, 37(3): 54-65.
[30] Xian Guojian, Zhao Ruixue, Zhu Liang, et al. SKOS transformation and application research of agricultural science thesaurus[J]. New Technology of Library and Information Service, 2012(10): 16-20.
[31] Xian Guojian, Zhao Ruixue, Kou Yuantao, et al. Research and practice on agricultural science thesaurus linked data construction[J]. New Technology of Library and Information Service, 2013(11): 8-14.
[32] Liu Huamei. Research on SKOS description and automatic conversion of subject terms in Chinese Classification Thesaurus[J]. Library Development, 2014(8): 29-32, 36.
[33] Ou Shiyan. Semantic conversion of Chinese thesauri[J]. Library and Information Service, 2015, 59(16): 110-118.
[34] Radulovic F, Poveda-Villalón M, Daniel Vila-Suero, et al. Guidelines for Linked Data generation and publication: an example in building energy consumption[J]. Automation in Construction, 2015, 57: 178-187.
[35] Xia Cuijuan, Liu Wei, Zhao Liang, et al. Linked data publication technology and implementation: taking Drupal as an example[J]. Journal of Library Science in China, 2012, 38(1): 49-57.
[36] Shen Zhihong. Research on linked data publication process and key issues: taking scientific literature and data publication as examples[J]. Journal of Library Science in China, 2013, 39(2): 53-65.
[37] Fan Wei, Zou Qing. Exploration of term network services for Chinese Classification Thesaurus: taking subject term normative data as an example[J]. Library and Information Service, 2012, 56(14): 40-48.
[38] Hong Na, Qian Qing, Fan Wei, et al. Visualization practice of relationship discovery in linked data[J]. New Technology of Library and Information Service, 2013(2): 11-17.
[39] Hong Na, Zhu Kai, Wang Junhui, et al. Implementing biomedical semantic relationship discovery using RelFinder[J]. Journal of Intelligence, 2013(4): 142-148.
[40] Ren Ruijuan, Pu Demin, Zhang Yuan. Visualization practice of knowledge context based on five-dimensional academic relationship discovery[J]. Journal

of Academic Libraries, 2016(1): 69-75.
[41] Shi Zeshun, Xiao Ming. Practice of discovering semantic relationships in LIS linked data based on RelFinder[J]. Library and Information Service, 2017, 61(17): 139-148.
[42] Zhao Longwen, Luo Lishu. Government data opening based on linked data: models, methods, and implementation[J]. Library and Information Service, 2017, 61(19): 102-112.
[43] Chen Tao, Xia Cuijuan, Liu Wei, et al. Research and implementation of linked data visualization technology[J]. Library and Information Service, 2015, 59(17): 113-119.
[44] Liang A C, Sini M. Mapping AGROVOC and the Chinese agricultural thesaurus: definitions, tools, procedures[J]. New review of hypermedia and multimedia, 2006, 12(1): 51-62.
[45] Morshed A, Caracciolo C, Johannesen G, et al. Thesaurus alignment for linked data publishing[C]//Proceedings of the 2011 international conference on Dublin core and metadata applications. Hague: The National Library of the Netherlands, 2011.
[46] Zhu Wenjing, Xia Cuijuan, Liu Wei. Comprehensive analysis of SILK linking discovery framework[J]. New Technology of Library and Information Service, 2013(4): 18-24.
[47] Ou Shiyan, Hu Shan, Zhang Shuai. Ontology and linked data-driven semantic integration methods for library information resources and their evaluation[J]. Library and Information Service, 2014, 58(2): 5-13.
[48] Yu Wei, Chen Junpeng. Research on bibliographic data linking matching based on MapReduce[J]. New Technology of Library and Information Service, 2013(9): 15-22.
[49] Zhong Yuanxin, Li Tianzhang, Liu Wei. Research on OPAC mashup linked data application[J]. New Technology of Library and Information Service, 2013(4): 25-29.
[50] Ou Shiyan. Design and application of terminology registry and service system based on SOA architecture[J]. Journal of Library Science in China, 2011, 37(5): 13-25.
[51] Ou Shiyan, Tang Zhengui, Su Feifei. Research on terminology service construction and application for information retrieval[J]. Journal of Library Science in China, 2016, 42(2): 32-51.
[52] Zeng Xinhong, Huang Huajun, Liu Chunyan, et al. Research on ISO 5127 SKOS semantic description scheme and its shared service system[J]. Library and Information Service, 2017, 61(21): 123-129.
[53] Huang Huajun, Zeng Xinhong, Lin Weiming, et al. Research on formal semantic description standard system for Chinese knowledge organization systems (II): research and implementation of classification shared service system CLSS[J]. Journal of Library Science in China, 2015, 41(2): 17-28.
[54] Xu Lei, Dong Hui. Implementation of terminology registry and service based on REST architecture[J]. New Technology of Library and Information Service, 2012(7/8): 59-65.
[55] Wang Jun. Automatic ontology construction based on traditional knowl-

edge organization resources[J]. Journal of the China Society for Scientific and Technical Information, 2009, 28(5): 651-657.
[56] Ou Shiyan, Tang Zhengui. Research on automatic question answering technology for library linked data[J]. Journal of Library Science in China, 2015, 41(6): 44-60.
[57] Chang Jin. Compilation and development of thesauri in network environment[M]. Beijing: Science and Technology Literature Press, 2015.
[58] Sun Tan, Liu Zheng. Thoughts on constructing knowledge organization system for foreign scientific and technical literature information[J]. Journal of the China Society for Scientific and Technical Information, 2013, 32(7): 730-738.
[59] Liu Zheng, Ji Shanshan. Research on data model of thesaurus standards[J]. Library and Information Service, 2013, 57(2): 103-108.
[60] Wu Wenna, Bao Xiulin. Architecture and data model of national thesaurus database[J]. Journal of Library Science in China, 2016, 42(2): 81-96.
[61] Bao Xiulin, Wu Wenna. Analysis of semantic mapping quality and influencing factors[J]. Journal of Library Science in China, 2016, 42(5): 57-67.
[62] Binding C, Tudhope D. Improving interoperability using vocabulary linked data[J]. International journal of digital libraries, 2016, 17(1): 5-21.
[63] Zeng M L. Smart data for digital humanities[J]. Journal of data and information science, 2017, 2(1): 1-12.
[64] Allen B P. The role of metadata in the second machine age[EB/OL]. [2018-08-10]. http://dcevents.dublincore.org/IntConf/dc-2016/paper/view/464/534.
[65] Li Fang, Liu Shengyu, Liu Zheng. Review of biomedical semantic relationship extraction methods[J]. Library Tribune, 2017(6): 61-69.
[66] Choudhury S. The role of metadata in an open knowledge age?[EB/OL]. [2018-08-10]. http://dcevents.dublincore.org/IntConf/dc-2017/paper/viewFile/526/648.
[67] Wang Jun, Zhang Lu, Zhang Wenjun. Design of e-commerce shopping guide mechanism based on user needs[J]. Journal of Library Science in China, 2016, 42(5): 57-67.
[68] Gracyk F, Zeng M L, Skirvin L. Exploring methods to improve access to music resources by aligning library data with linked data: a report of methodologies and preliminary findings[J]. Journal of the American Society for Information Science and Technology, 2013, 64(10): 2078-2099.

Abstract: [Purpose/significance] Semantic organization of vocabulary, an important part in collection semantic research, is the focus of knowledge organization study. A research review in this field is helpful to promote its development. [Method/process] Based on the analysis of core terms in the field of vocabulary semantic organization, this paper proposes the analytical framework of "standard specification—semantic organization method—supporting technology—vocabulary application". With above framework, the paper reviews literature about method, technology and application. [Result/conclusion] Firstly, the paper gives the definition and main frame of vocabulary semantic organization, discusses the core concepts and their relationship including

vocabulary, ontology and linked data. Then taking the example of thesaurus, it summarizes the typical research of vocabulary semantic organization in China in recent ten years. And it compares the traditional vocabulary research and semantic research. On the basis of summarizing the above literature, the current situation and future development of semantic organization of Chinese vocabulary are discussed.

Keywords: vocabulary, linked data, semantic Web, resource descriptive framework

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