Research on the Application and Development of Thesauri

Shuai Yuanhua[1] and Zhao Jing[2]

[1]School of Software and Microelectronics, Peking University
No. 24, Jinyuan Road, Daxing District, Beijing, 102600, China
shuaiyuanhua@163.com

[2]Institute of Scientific and Technical Information of China
No.15 Fuxing Road, Haidian District, Beijing, 100038, China

ABSTRACT. *A thesaurus serves as an important tool mainly for subject indexing and information retrieval. This paper introduces some typical thesauri and relevant researches on their development and application. Through studying and researching existing online thesaurus platforms, this paper analyzes these platforms' function and application. Based on the study above, this paper puts forward the application and development trend for thesauri.*
Keywords: thesaurus, application, development

1. Introduction. A thesaurus is a controlled and structured vocabulary in which concepts are represented by terms, organized so that relationships between concepts are made explicit, and preferred terms are accompanied by lead-in entries for synonyms or quasi-synonyms.[i] As a semantic dictionary, it's composed of terms in professional fields and semantic relationship between terms. Thesauri are intended to support information retrieval and data indexing as well as reduce users' professional knowledge requirements when searching.

Soergel[ii] divides thesauri into two kinds — term-based thesauri and concept-based thesauri. The latter mainly consist of descriptors, non-descriptors, term relationships and category index. Concept-based thesauri take concepts as core which are described by terms, of which preferred terms are used to represent concept while non-preferred terms are regarded as entry terms. Term relationships are composed of hierarchical relationship,

equivalence relationship and associative relationship. The hierarchical relationship presents the broader and narrower term of a term; the equivalence relationship connects a term's synonyms; and the associative relationship means relationship between a pair of concepts which are not hierarchically related but share a strong semantic connection.

Currently thesauri are mostly based on the Internet rather than paper. Information technology promotes thesauri to become electronic tools gradually, which help organize knowledge by integrating thesaurus into search engines or content management software. Online thesauri support precise knowledge description and knowledge retrieval on different information platforms meanwhile increasing the availability of information. Therefore, it's essential to explore and research the existent thesauri and their studies, use online thesaurus platforms and analyze their function and application.

2. **Related Work.** The construction of the network thesaurus started earlier in foreign countries. According to the Taxonomy Warehouse website[iii], there are about 700 taxonomies, thesauri, glossary, and ontology used in the network environment in Europe and America. They are mainly used for the organization and optimization of network information resources such as network databases, search engines and Web pages. There are over 40 bilingual or multilingual thesauri. The typical thesauri include the agricultural thesaurus AGROVOC used by the FAO database, the National Medical Library's Medical Thesaurus (MeSH), the European Multilingual Thesaurus (EuroVoc) and the US Library of Congress Subject Headings (LCSH) and so on.

Foreign countries attach great importance to the construction of multi-language thesaurus. The research work of building and application of multi-language thesaurus is emerging a lot. The "MACS Project"[iv], funded by the Conference of European National Librarians, maps the RAMEAU (Répertoire d'autorité-matière encyclopédique et alphabétique unifié) with LCSH and GND (Integrated Authority File) to obtain the catalog of multilingual books using the existing indexing tools. A.C. Liang[v] adapts the methods of mapping to integrate the English AGROVOC and Chinese agricultural thesaurus. J. Ramírez et al.[vi] take English as the intermediate language and use Wikipedia and WordNet to establish multi-language thesauri including English, Japanese and Spanish. Prof. C. C. Yang of the Chinese University of Hong Kong used the Hopfield neural network[vii] and the associative constraint network[viii] to construct bilingual subject headings based on English and Chinese parallel corpus. Zeng Wen et al.[ix] explore the multi-language thesaurus automatic construction methods and related technologies. Xu Hongjiao et al.[x] research the translation of multi-language thesaurus based on analyzing the shortcomings of translation, and then from four aspects to localize English EI thesaurus and the Japanese JST thesaurus to Chinese.

The construction of ontology from thesauri and the visualization of thesauri are hot research spots. Roberto Navigli and Paola Velardi[xi] describe a pattern-based methodology to automatically enrich a core ontology with a thesaurus and other controlled vocabularies. Daniel KLESS et al.[xii] present a method that provides clear steps to follow when re-engineering a thesaurus into ontology. The method makes use of top-level ontologies

and was derived from the structural differences between thesauri and ontologies as well as from best practices in modeling. Li Hua et al[xiii] use the graphical display function of Protégé to visualize the Chinese traditional tool professional thesaurus. Jia Junzhi[xiv] proposes the SKOS (Simple Knowledge Organization System) representation method of Chinese Thesaurus".

Saif Mohammad et al.[xv] propose an approach to generate a high-coverage semantic orientation lexicon, which includes both individual words and multi-word expressions, using only a Roget-like thesaurus and some affixes. Dolf Trieschnigg et al.[xvi] compare six MeSH classification systems to find A KNN system outperforms other approaches. They draw a conclusion that the annotation of biomedical texts using controlled vocabularies such as MeSH can be automated to improve text-only IR. Shanfeng Zhu[xvii] proposes a new strategy to apply the semantic information of MeSH thesaurus to cluster MEDLINE documents through mapping documents into MeSH concept to be clustered. Wu Dongmin and Xu Yiwei[xviii] summarize the characteristics of thesauri under the network environment, and give advice to the maintenance of thesaurus.

3. **Application of Thesauri.** With the rapid increase of online information resources, information retrieval has become not comprehensive and accurate. As for this issue, thesauri, forgot by researchers at one time, has once again attracted researcher's focus. Researchers use thesauri to solve the issue that information retrieval is not comprehensive and accurate. The reason is that thesauri have advantages such as high indexing accuracy, depth and specificity and it's easy to use. And they play an important part in information retrieval through multiple ways and factors, which greatly improves retrieval and work efficiency.

In the past, thesauri were designed for information professionals trained in indexing and searching. At present there is a huge demand for vocabularies that untrained users find to be intuitive, and for vocabularies that enable inferencing by machines. This demand results in the appearance of electronic thesauri, which have been applied to many fields, such as knowledge navigation, information classification, data mining, semantic retrieval and so on. AGROVOC, a thesaurus developed and maintained by the Food and Agriculture Organization of the United Nations, contains 21 languages and has been applied in agriculture comprehensively. American NASA thesaurus has made many achievements in aerospace. Big data call for new demand of application, not only in information agencies but also relevant cross-domain agencies. For example, nowadays developed countries have applied thesauri in description and extraction of online government information. Researchers have used thesauri to construct automatic classification system of repository and to realize intelligent reason and semantic clustering of concept relationship. Besides, thesauri have been applied in knowledge management of e-commerce and large enterprises, where they are used to realize automatic indexing of mass information. Thus we can realize intelligent inquiry, domain monitoring, knowledge mining and domain knowledge clustering under different granularity. As for domains of medicine and education, thesauri have been also been used in word segmentation, information extraction and clustering,

word frequency count, sentiment analysis and other fundamental text information processing. And by indexing bilingual information with thesauri, we can develop cross-language retrieval and assistant machine translation.

Thesauri also exert great influence after they are shifted to ontology. Decades of development and improvement have made thesauri relevantly mature in structure. By absorbing the concepts, terms and structure of thesauri, we can accelerate the shift from thesauri to ontology. And this shift can expand the application scope of thesauri, such as knowledge navigation, information classification, data mining and semantic retrieval.

4. **Thesaurus Platforms.**

4.1. **UMLS.** UMLS (United Medical Language System) is a knowledge organization system containing biomedical vocabularies built up by the US National Library of Medicine. The UMLS integrates over 2 million names for about 900 000 concepts from more than 60 families of biomedical vocabularies, as well as 12 million relations among these concepts.[xix]The UMLS consists of four parts: Metathesaurus, Semantic Network, SPECIALIST Lexicon and other supporting tools. Among these parts Metathesaurus is the most-used knowledge source, succeeded by Semantic Network and SPECIALIST Lexicon. The UMLS has been widely applied in information retrieval, natural language processing, electronic patient records, health data standards and so on.[xx]According to UMLS User Annual Reports, its application mainly includes terminology research, terminology mapping, creating local terminology and other application including indexing, retrieval and natural language processing.

4.1.1. **Metathesaurus**. Metathesaurus, the core of UMLS knowledge sources, is composed of concepts and terminologies derived from varieties of controlled vocabularies as well as their relationships. The fundamental unit in Metathesaurus is concept, which has different expressions in different vocabularies. The UMLS Metathesaurus organizes terms to build a concept by concentrating different description of same thing. A more popular term is adapted for preferred term to describe a concept, and a concept unique identifier is given to this concept. Synonyms of the popular term are also given a lexical unique identifier. Therefore, the organization model of UMLS Metathesaurus is strings-terms-concepts. The Metathesaurus is widely used as the underlying source for dictionary-based natural language processing (NLP) systems, such as MetaMap for recognizing biomedical concept and SemRep for extracting relationship from the biomedical literature and from clinical documents.[xxi]

4.1.2. **Semantic Network.** The UMLS Semantic Network connects all the concepts of thesauri closely. It is divided into two parts: semantic types and semantic relationships. The former one is a group of broad categories which offer a unified category for all the concepts in Metathesaurus. So far semantic types contains two major types — entity and event. Semantic relationships refer to a group of relationships existing among semantic types. There are 54 kinds of semantic relationships. Semantic types are the nodes in Semantic

Network while semantic relationships are connections between these nodes. NLM (National Library of Medicine) collects all concepts and clusters them to generate concept categories, namely semantic types. By building semantic relationship between semantic types, NLM can organize all the concepts under unified categories and hence it can reveal more semantic relationships than traditional thesauri. Semantic Network tries to make computers understand knowledge of general meaning, reason and make conclusion through semantic types and relationships. The semantic transits the semantic relationship from linear pattern to reticular pattern, exposing more complex and abundant semantic relationships between concepts.

4.1.3**. SPECIALIST Lexicon.** SPECIALIST Lexicon is a repository of dictionary information used for natural language processing. It collects common English words, terms of biomedicine as well as terms appeared in Medline, UMLS and Metathesaurus. Each term records dictionary information required by NLP system, including syntax, morphology and spelling. Besides, SPECIALIST Lexicon offers dictionary tool and application for Metathesaurus to determine the scope of English vocabulary and identify metaplasm of terms in texts and biomedicine. It's a powerful tool to retrieve, index and process words. Lexicon entries can be terms composed of a single word or multiple words. They include four parts: fundamental form, parts of speech, unique identifier and any readily available spelling forms.

SPECIALIST Lexicon offers some NLP tools. Lexical Access Tool supplies access to SPECIALIST Lexicon, and outputs TXT and XML files. SPECIALIST Text Tools help users to disintegrate free text into words, terms, phrases, sentences and paragraphs. Text Categorization can categorize and cluster in a higher level, which help segment texts, index content, retrieve records and disambiguate words and sentences. GSpell gives spelling advice, and dTagger is a tool for part-of-speech tagging.

4.1.4**. Other Supporting Tools.** UTS (UMLS Terminology Services) offers navigation, search and data acquisition for UMLS knowledge source by web browser and client. Its major tools include Metathesaurus browser, semantic web browser and SNOMED CT browser, which can help users search concepts, semantic types and relationships as well as content from SNOMED CT.

MetaMap is a widely used tool for mapping free texts to concepts in UMLS Metathesaurus, which means marking all the concepts in biomedical texts which appear in Metathesaurus. MetaMap is applied wildly in many aspects. For example, in the data retrieval of MEDLINE, it can improve literature retrieval results. And it is also applied in data mining, which includes clinical finding, and discovery of relationships between medicine and illness in literatures. Furthermore, MetaMap serves as a foundation for realizing NLM automatic indexing system, which is used for recommending terms for semi-automatic and full-automatic indexing.

SemRep, a semantic knowledge representation tool, applies NLP technology and the UMLS SPECIALIST Lexicon tool to segment sentences and tag part of speech of

biomedical texts. And it uses MetaMap to acquire concept, corresponding semantic types and relationships by mapping terms to concepts from Metathesaurus. The main argument of texts can be attained by concurrence of concepts, which presents in the form of subject-relation-object triples. Subject and object arguments of each predication are concepts from the UMLS Metathesaurus and the relation (in uppercase) is a relation from the UMLS Semantic Network.[xxii]

One of the SemRep's application is question-answering system. It's a specific method of information retrieval. Users input natural language to retrieve and the system outputs specific texts from literature by searching. The core technologies of question-answering system are texts processing and literature retrieval. In open field, to recognize named entities, such as names, places, organizations and time, is also important for this system.

4.2. **EuroVoc.** EuroVoc thesaurus is a multilingual and multi-domain vocabulary of the European Union. It aims to offer a unified indexing tool for information management and dissemination. At present, EuroVoc thesaurus covers all official languages of the member states of the European Union and other non-member countries. It includes 21 fields, 127 micro thesauri and 6833 descriptors. All concepts are accompanied by a reference to a micro thesaurus. Users can browse and search the latest version of thesauri, and download the EuroVoc thesauri according to alphabetical order or multiple languages.

EuroVoc thesaurus has many advantages. It builds up a standard for indexing vocabularies and allows users to search literature more accurately. The multiple language equivalence of concepts can bring great convenience for indexing and promotes translation.

4.3. **AGROVOC.** AGROVOC is a multilingual agricultural thesaurus developed by Food and Agriculture Organization (FAO) of United Nations, which covering not only agriculture but also food, fisheries, forestry and the environment. At present AGROVOC is composed of more than 30000 concepts available in 27 languages. FAO offers the navigation, retrieval and download of AGROVOC thesaurus, which is in format of MySQL, Microsoft Access, XML and OWL. AGROVOC is made available by FAO as an RDF/SKOS-XL concept scheme and published as a linked data set aligned to 16 other vocabularies.[xxiii]

As an important multilingual thesaurus, AGROVOC is widely used in organizing and retrieving information resources of agriculture domain. Users can use AGROVOC to find the equivalence of some plants in another language, or to find commodities and from which plant they are produced. AGRIS, one of the largest data base in the world, is organized by using AGROVOC. To encourage system developer of agricultural information to apply AGROVOC

4.4. **MeSH.** Medical Subject Headings (MeSH) is a large-scale comprehensive thesaurus built up and maintained by US National Library of Medicine (NLM). It has been translated into more than 20 languages including Chinese and widely used in biomedical field. MeSH was first published in 1960 and updated every week. Each year a new version of MeSH is

published. Since 2007, NLM has stopped publishing MeSH in paper. MeSH consists of two parts. The first part is an alphabetical list which is arrayed in the order of alphabet. The other part is a tree structure list, also named category list. Users can download subject headings, subheadings, supplementary concepts and tree structure in the format of XML, ASCII and MARC. MeSH supports web services to help users search headings and supplementary concepts, look up the whole records of MeSH and relationships between concepts. Its tool, MeSH on Demand, identifies MeSH terms in text using the NLM Medical Text Indexer (MTI) program.

After over 50 years of improvement and updates, the structure of MeSH tends to be more completed and sound so that it's adopted by information organizations all over the world. It's widely used to index and retrieve journals, books, audio-visual materials. MEDLINE database and Chinese biomedical literature database have used MeSH to index and retrieve journal articles. Furthermore, the MeSH Thesaurus has been incorporated for use in the PORTAL-DOORS System (PDS) for resource metadata management on the semantic web.[xxiv]

4.5. **Other Thesauri.** Online thesauri which support user interface mainly include browser pages, retrieve pages and pages for retrieved results. It has direct influence on users' emotion and actions whether web pages are easy to use, friendly and interactive. According to research on online thesaurus links on Taxonomy Warehouse, most online thesauri support keyword searching. And some thesauri such as AIATSIS thesaurus support forward matching and backward matching and substring searching. Thesauri like Health Thesaurus support precise retrieval methods including data, relationship between words, scope, languages and other conditions. Other thesauri like National Criminal Thesaurus support Boolean retrieval. Most thesauri support to look up vocabularies in alphabetical order. And some like Transportation Research Thesaurus allow users to browse according to subjects, hierarchy, KWIC and KWOC. But some thesauri only support alphabetical browsing such as Federal Register Thesaurus of Indexing Terms.

Some thesauri support both browsing and retrieving, such as British Education Thesaurus. Its homepage includes keyword retrieval, alphabetical browsing, search screening and search history. CAB Thesaurus supports precise searching and browsing by alphabetic and subjects. Art and Architecture Thesaurus allows keyword searching and hierarchy browsing. But some thesauri only support searching without browsing service, such as Bioethics Thesaurus and UNICRI Thesaurus. Asian Vegetable Thesaurus supplies foreign trader with searching and retrieving services aiming to help the export of Asian vegetables. Thesauri like this don't have complex inquiry modes, and they presents acknowledge intuitively with pictures.

Almost every online thesaurus has built hyperlinks between words, only a little are in static HTML and TXT formats. Furthermore, some thesauri allow users to download thesauri in PDF, XML and SKOS/RDF formats as well as suggest revising terms through email. African Studies Thesauri, Getty Thesaurus of Geographic Names and so on offer web services such frequently asked question. General Multilingual Environmental

Thesaurus and STW Thesaurus for Economies support Wiki collaboration. Health Thesaurus and General Multilingual Environmental Thesaurus offer RSS service.

5. **Development Trends of Thesauri.** The development of thesauri has experienced a period of low ebb and tardiness, but currently it ushered in new development opportunities. The issue of ISO 25964, the international standard for thesauri, promotes the formulation of other countries' thesauri standards such as China's GB/T 13190.1-2015. These new standards for thesauri give directions for traditional thesauri's to adapt to network knowledge organization and to construct new thesauri.

Currently in developed countries, thesauri have been widely used in medicine, education, agriculture, business and industry as well as government, commodity and knowledge management. The application of thesauri has formed a large scale. AGROVOC Thesaurus has been adapted into RDF/SKOS-XL. UMLS Metathesaurus, ERIC Thesaurus and Humanities and Social Science Electronic Thesauri have already adapted themselves to the requirements of information indexing and retrieval. Some countries apply thesaurus in specific careers and professional domains. For example, Thesaurus of Australian Government Subjects and Australian Governments' Interactive Functions Thesaurus are used to describe government information and offer information retrieval. For some developing countries, the application scope of thesauri should extend from libraries to all walks of life. The goal of socialization for thesauri is to make them become a universal tool for knowledge organization, support precise knowledge description and retrieval on all kinds of application platforms as well as improve the availability of platform information.

The challenges that thesaurus face promote it to develop in two main directions: one is to make thesauri easier, more convenient and intuitive to use; the other is collaboration between systems, which means thesaurus can be embedded into catalogue management systems, search engines and subordinate application systems of web portal. In the web era, the thesaurus, drew up and revised based on characteristics of network information, is still an important tool for information and knowledge organization. Its main application scope includes information indexing and retrieval, subject classification navigation, machine translation, semantic network and word segmentation. Also it connects closely to standard data protocol, mapping or interoperability, visual user interface, subject map, transition to ontology and so on. To conclude from studies and research above, the development and application trends of thesauri can be classified into aspects as follows:

5.1. **Multilingual thesauri.** With the increasing demand for information from different cultural and linguistic backgrounds, multilingual retrieval tools become more and more important, and multilingual thesaurus has become a trend in the compilation and development of thesaurus. Thesaurus is usually not a multi-language version at the beginning, but with the information needs of organizations, it is added a new language version by translating the original thesaurus in another language.

5.2. **Interoperability with other vocabularies and thesauri.** The interoperability includes

the interoperability of thesauri with the controlled vocabulary, as well as the mapping of different thesauri. For example, in addressing the issue of vocabulary interoperability, the Australian Thesaurus of Education Descriptors allows users to link to the British Education Thesaurus, the Canadian Literary Thesaurus, the EU Education Thesaurus, the School Online Thesaurus and the Subjects of New Zealand by clicking the Link button on its website.

5.3. **Visualized thesauri.** Thesauri should be easy for users to understand by improving the precision and recall of user retrieval with definitions, scope notes, or introductory paragraphs. Under the network environment, the construction and application of thesauri can adapt the corresponding technology to achieve visualization of the relationships between words, which hence improves the convenience of browsing and searching. For example, the Protégé tool, developed by the Medicine and Informatics Research Group at Stanford University School of Medicine, can be used to visualize relationships between words. And the Visual Thesaurus uses the radiation pattern to display related words and allows users to click on the query words.

5.4. **Intelligent update mode.** Network thesaurus is based on the concept modeling so that it can fully realize the automatic construction, automatic enrichment and updating, automatic indexing and knowledge discovery. For example, the MeSH automatically replaces the old vocabulary with new vocabulary by co-word clustering analysis. The essence of the method is to extract the new words from the indexed corpus and then analyze the relationship between them and the indexed words so as to determine the standard words of new words. Finally the new words are added to the vocabulary as the narrower terms of the corresponding terms. Therefore the network thesaurus achieves automatic updates.

5.5. **Transition to ontology from thesauri.** The transition from network thesaurus to ontology can both maintain its continuity and inheritance but also its semantic network environment. It's one of the development direction under semantic web. There are many similarities between thesauri and ontology. Both provide a set of frameworks for the concepts and terminologies established for disciplines or in natural language. Thesauri and ontology provide vocabulary and structural relationships between words, but the latter provides semantic relationships and other information about concepts and their relationships. So the ontology can provide better knowledge representation and powerful reasoning.

5.6. **Application of Web 2.0.** At present, under the Web 2.0 environment with applications such as Blog, SNS, RSS, Wiki and so on, user participation and interaction, collective wisdom, co-construction and sharing are important characteristics of Web2.0. Web thesauri also begin to introduce Web2.0 technology for vocabulary update and maintenance as well as interaction with users. In the context of Web collaboration, the construction of the network thesaurus needs to transit from organization-oriented information indexing to

user's information retrieval, from single service to multi-purpose service and provide RSS, BLOG and other personalized services. To make full use of the users' collective wisdom, methods like Wiki can be used to build thesaurus collaboration platform, where users can participate in the common construction or maintenance of thesaurus as well as vocabulary optimization.

5.7. **Strengthening the indexing and retrieval application of thesauri.** Using thesauri for indexing and retrieval can improve the precision and recall of search. Thesaurus can be directly connected with the search engines, databases through the network, which brings great convenience for users to search. Most large databases, in order to ensure the retrieved results accurate and comprehensive, still use thesauri to index, which making the thesauri get maintained and updated.

5.8. **Intuitive thesaurus display and operation.** First, as for the main search page and the display of the search entry word, use "recommended vocabulary", "subject headings", "descriptors" and so on. Provide thesaurus browsing by category, alphabet and classification, multiple retrieval methods and entrances and support browsing and retrieval strategies. Second, provide a simple, understandable term to describing relationships between identifiers and vocabularies, such as "NT", "BT", "RT", "USE" and other symbols, or use complete forms to describe the relationships between the thesaurus, such as "narrower term", "broader term" "related term", etc., and may use "+","-","•"or the like. Third, for the specific mode of operation, provide functions like dragging, hypertext links, double-clicking and other flexible ways. For example, words can be directly pasted in the database retrieval system so as to save user time. To store the words chosen when users browse the thesaurus, web platform can provide "vocabulary storage" function. Fourth, as for the access to resources, online thesaurus platform can offer resource links directly under the relevant words. Retrieval interface is developing in the direction of individualization, such as the National Monuments Record Thesauri, which supplies new users and regular users with two different interfaces respectively. The display and operation of vocabulary should be easy and convenient. Help file is particularly important for new users. The help file should include the menu introduction, topic classification, FAQ, quick search and so on.

6. **Conclusion.** This paper introduces some researches on the development and application of thesauri. Through exploring some typical thesauri and their online versions which offer Web services, this paper analyzes their application. Based on these researches, this paper puts forward the application and development trend of thesauri — multilingual, interoperable, visualized, intelligible and individualized.

REFERENCES

[1] International Organization for Standards. ISO 25964-1 Information and Documentation-Thesauri and Interoperability with Other Vocabularies-Part 1: Thesauri for Information Retrieval[S]. Geneva: International Organization for Standardization, 2011.

[2] Soergel, Dagobert. Data Models for an Integrated Thesaurus Database. (2001).

[3] Taxonomy Warehouse. Thesaurus [EB/OL]. [2016-12-26]. http://www.taxonomywarehouse.com/

[4] MACS [EB/OL]. [2016-12-26]. http://www.dnb.de/EN/Wir/Kooperation/MACS/macs.html

[5] A. C. Liang, and M. Sini. Mapping AGROVOC and the Chinese Agricultural Thesaurus: Definitions, tools, procedures. *New Review in Hypermedia & Multimedia* 12.1(2006):51-62.

[6] Ramírez, Jessica, M. Asahara, and Y. Matsumoto. Japanese-Spanish Thesaurus Construction Using English as a Pivot. *Eprint Arxiv* (2013).

[7] Yang, Christopher C., and J. Luk. Automatic generation of English/Chinese thesaurus based on a parallel corpus in law. *Journal of the Association for Information Science and Technology* 54.7(2003):671-682.

[8] Yang, Christopher C., C. P. Wei, and K. W. Li. Cross-lingual thesaurus for multilingual knowledge management. *Decision Support Systems* 45.3(2008):596-605.

[9] Zeng Wen, Wang Huilin. Research on Automatic Construction of Cross-lingual Thesaurus. *Library and Information Work* 55.4 (2011): 106-109.

[10] Xu Hongjiao, et al. Research and Practice on the Construction of Multi-language Thesauri. *Library and Information Service* 19 (2014): 7-12.

[11] Navigli, Roberto, and P. Velardi. Enriching a formal ontology with a thesaurus: an application in the cultural heritage domain. *COLING• ACL 2006* (2006):1-9.

[12] Kless, Daniel, et al. A method for re-engineering a thesaurus into an ontology. *International Conference* 2012.

[13] Li, Hua, J. Liu, and H. Hou. The Visualization of Thesaurus by Using an Ontology Editing Tool-Protégé. *Library Journal* (2007).

[14] Jia Junzhi, Yang Jie, Wei Rongjun. Automatic Transformation Design of Simple Knowledge Organization System Representation of Chinese Subject Headings [J] .*Journal of Intelligence Theory and Practice*, 2011,34

[15] Mohammad, Saif, Cody Dunne, and Bonnie Dorr. Generating high-coverage semantic orientation lexicons from overtly marked words and a thesaurus. *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 2-Volume 2*. Association for Computational Linguistics, 2009.

[16] Trieschnigg, D, et al. MeSH Up: effective MeSH text classification for improved document retrieval. *Bioinformatics* 25.11(2009):1412-1418.

[17] Zhu, Shanfeng, J. Zeng, and H. Mamitsuka. Enhancing MEDLINE document clustering by incorporating MeSH semantic similarity. *Bioinformatics* 25.15(2009):1944-1951.

[18] Wu Dongmin, and Xu Yiwei. Research on the Development of Subject Headings in Network Environment. *Library Science Research* 4 (2016): 43-48.

[19] Bodenreider, Olivier. The unified medical language system (UMLS): integrating biomedical terminology. *Nucleic acids research* 32.suppl 1 (2004): D267-D270.

[20] UMLS Applications. [EB/OL]. [2016-12-28]. https://www.nlm.nih.gov/research/umls/implementation_resources/applications.html

[21] Xu, R., M. A. Musen, and N. H. Shah. A Comprehensive Analysis of Five Million UMLS Metathesaurus Terms Using Eighteen Million MEDLINE Citations. *AMIA. Annual Symposium proceedings / AMIA Symposium. AMIA Symposium* 2010(2010):907-11.

[22] National Library of Medicine. SemRep [EB/OL]. [2016-12-26]. https://semrep.nlm.nih.gov/

[23] Food and Agriculture Organization. AGROVOC [EB/OL]. [2016-12-26]. http://aims.fao.org/vest-registry/vocabularies/agrovoc-multilingual-agricultural-thesaurus

[24] Taswell, C. Use of NLM medical subject headings with the MeSH2010 thesaurus in the PORTAL-DOORS system. *Studies in Health Technology & Informatics* 159(2010):255-8.