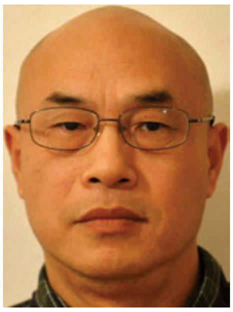
Construction and Evolution of a Chinese Information Science and Information Service (CIS&IS) Onto-Thesaurus†

Zhao Youlin*, José Miguel Baptista Nunes**, Deng Zhonghua***

*School of Information Management in Wuhan Unversity, Luo-jia-shan, Wuchang, Wuhan, Hubei Province, P.R.China, 430072, <simzyl@whu.edu.cn>
**Information School University of Sheffield, Regent Court, Sheffield S1 4DP, <j.m.nunes@sheffield.ac.uk>
***School of Information Management in Wuhan Unversity, Luo-jia-shan, Wuchang, Wuhan, Hubei Province, P.R.China, 430072, <sim.dzh@gmail.com>

Zhao Youlin is a Ph.D. candidate at the School of Information Management, Wuhan University, China. She has published many research papers in journals such as *Journal of the China Society for Scientific and Technical Information* and *Library and Information Service,* and has had international conference papers indexed by EI compendex. Her research interests include knowledge organization and knowledge management, information analysis and information processing, digital library and cloud computing service.

Miguel Baptista Nunes, BSc, MSc, Ph.D., FHEA, FBCS, is a Senior Lecturer in Information Management at the Information School, University of Sheffield. Miguel is currently the Head of the Information Systems Research Group in the School and has been involved in research in information management and information systems for the last 20 years. He has published more than 150 refereed articles in both academic conferences and academic journals, published a book on action research for e-learning and served as programme chair for a number of International conferences.



Deng Zhonghua is a professor in School of Information Management at Wuhan University. He received his Master's Degree from Wuhan University of Hydraulics and Electricity in management information system in 1993 and his Ph.D. from the same university in hydrology and water resources in 2000. His principal research interests lie in the fields of knowledge organizing and processing, information services and information management around library science.

Youlin, Zhao, Baptista Nunes, José Miguel, and Zhonghua, Deng. **Construction and Evolution of a Chinese Information Science and Information Service (CIS&IS) Onto-Thesaurus.** *Knowledge Organization. 41(2)*, 131-144. 51 references.

Abstract: Thesauri are the most important tools for information and knowledge organization, and they undergo regular improvements according to the rapid development of new requirements and affordances of emerging information techniques. This paper attempts to integrate ontology into the conceptual organization scheme of thesauri and proposes a new solution to extend the functionality of thesauri based on ontological features, which is termed here as an onto-thesaurus. In this study, a prototype system named the Chinese Information Science and Information Service onto-thesaurus system (CIS&IS), was developed to analyze onto-thesaurus with the category of *information science and information service* in the *Chinese Topic Classification Dictionary* with a two-stage approach. The first stage aims to define and construct the onto-thesaurus. The second stage aims to realize the evolution function of onto-thesaurus. The main purpose of this system was to achieve the function of self-learning and auto-evolution and to enable a much more effective conceptual retrieval by the newly proposed onto-thesaurus.

† This work was partially supported by National Natural Science Foundation of China under Grant NSFC-71373191 and by National Natural Science Foundation of China under Grant NSFC-71173163.

Received 19 April 2013; Revised 14 January 2014; Accepted 21 January 2014

Keywords: relationships, thesauri, information, onto-thesaurus, ontology, research, Chinese

1.0 Introduction

Knowledge organization and representation are crucial prerequisites for knowledge sharing and also important foundations for knowledge management. Thesauri, as a method of organizing and representing knowledge, use standard language and specific terms to characterize well defined subjects and domains. In general, a thesaurus is a reference tool that lists words grouped together according to similarity of meaning (containing synonyms and sometimes antonyms). Whilst having been widely used with very high degrees of success, the weaknesses of thesauri have become more apparent with the very rapid development of information technology (Losee 2007), namely:

- Long revision circles;
- High update cost;
- Severe lagging of constantly emerging neologisms;
- Lack of automatic updating capacity.

The discussion above indicates that there is a clear need to improve thesauri capacities in order to make the most of IT affordances today. This paper proposes that this can be achieve by associating characteristics of ontologies, as used in the semantic web, and other metadata enabled applications.

Computational ontologies are new artifacts for knowledge record and processing and also one of the foundations of semantic web (Marcondes 2013). The term ontology was initially connoted with the philosophical study of the nature of being, becoming, existence or reality, as well as the basic categories of being and their relations (Gruber 2009). Ontology originates from early Greek philosophy and aims at characterizing the nature of "being" through the entities that surround an existent object. Parmenides of Elea was among the first to propose an ontological characterization of the fundamental nature of existence; that is, the systematic study of the "existence" and "its essence" (Antoniou and Van Harmelen 2004). With progress in artificial intelligence, ontology has developed in its connotations. Ontology, as "an explicit formal specification of the terms in the domain and relations among them" (Gruber 1993) has become a widely accepted and commonly used conceptualization, mainly in information and computer sciences. As such, ontology is a term that is widely misused and as acquired meaning associated with models that use classes/objects to represent concepts or entities in a certain field (Kalinichenko et al. 2003). This type of model uses identified properties to qualify and reflect attributes or values that characterize the class/object. Furthermore, ontologies use relationships to represent the connection between the classes and objects (Bagni et al. 2007). The definition of relationships in the ontological model is different from the thesauri. The relationships of ontology are characterized by flexibility and the ability to expand.

The contents and focus of thesauri and ontologies are therefore not the same particularly in the aspect of vocabulary organization. Ontologies do not use standardized language as in thesauri. Ontology terms also include natural and semi-natural language. Furthermore, ontologies usually rely on machine indexing and automatic cataloguing whereas thesauri rely on manual indexing (White, Willis and Greenberg 2013, Willis and Losee 2013). In terms of accuracy ontologies are not as good as manual indexing methods (Gilchrist 2003). However, generally, ontologies have two significant advantages (Rosati 2005). A first advantage is the open and extendible nature of the established relationships. A second advantage is the evolution function.

This paper reports on an experimental study that aimed at designing and developing an onto-thesaurus evolutionary system that combines the very useful features of ontologies with the ones of basic thesauri. The context of the research is Chinese and deals with very specific cultural, terminological and representational aspects. Furthermore, the current body of knowledge in the vast Chinese research in this area is mostly concerned with the building of ontologies by adding thesauri features (Tang 2004, Tang et al. 2005, Zhang and Yu 2006, Sun and Wan 2007 and Yang and Sun 2013). In this research we are proposing that more attention should be given to extending the functionality of thesauri based on useful ontological features. Furthermore, there is in Chinese research a disconnection between theoretical ontology building and retrieval technology systems using ontologies, which makes it very often difficult to assess some of the theoretical propositions. This paper offers both theoretical propositions and their empirical evaluation with a Chinese academic journal corpus.

2.0 Literature Review of Related Work

2.1 Thesauri and Ontologies

There are thirteen kinds of relationships between vocabularies in the ANSI Thesauri Standard (Z39.19-2005) which contains specific structures ("Y", "D", "S", "F", "C" and "Z") for *Chinese Thesauri*. This enables it to achieve considerably better performance than others on semantic representations. Furthermore, Pastor-Sanchez et al. (2009) proved that the Simple Knowledge Organization System (SKOS) was one of the best options for thesauri representation on the Semantic Web.

However, there are alternatives to SKOS that may present some advantages. Ontology languages allow users to write explicit, formal conceptualizations of domains models. The main requirements are: (a) a well-defined syntax; (b) a well-defined semantics; (c) efficient reasoning support; (d) sufficient expressive power; and (e) convenience of expression (Antoniou and Van Harmelen 2009).

The Web Ontology Language (OWL) is such a semantic markup language for publishing and sharing ontologies on the Web. OWL is developed as a vocabulary extension of Resource Description Framework (RDF) (Bechhofer et al. 2004). OWL is very commonly used and has some significant advantages, such as model theoretic formal semantics based on strong logical foundations (Hitzler et al. 2009). Furthermore, it shares many characteristics with RDF (Horrocks and Patel-Schneider 2004).

OWL and SKOS are both encoding languages but have some significant differences. OWL is mostly used to resolve major issues related to classification systems and thesauri (Zeng, Panzer and Salaba 2010). SKOS is used in the development of specifications and standards to support the use of a knowledge organization system (KOS) with the framework of the Semantic Web. Therefore, SKOS provides a standard way to represent KOS's using RDF and, in the context of this research, uses a Chinese KOS. SKOS uses low formalization requirements, providing representational language and an API facility for its application on the Web. OWL has also been contextualized for Chinese through a DL-based onto-thesaurus that has specifically been developed for Chinese thesauri or any other Chinese KOS with high formalization requirements (Panzer and Zeng 2009).

Kless et al. (2012) discussed the differences and similarities between ontology and thesauri, and their findings made clear that thesauri require structural and definitional reengineering in order to be reused or treated as ontologies, but that adherence to the international standard for thesauri provides a good base for such reengineering. At first sight, the functional features of thesauri are covered by ontology, which can cause confusion, in that ontology is the developmental direction of thesauri. However, this is not the case. First of all, from a practical perspective, thesauri are recognized, but the application of ontology is less extensive than expected. The reason for is that building ontology requires much time, effort and expense. The explanation is obviously insufficient, because if ontology could have better effects than thesauri, we need to consider whether it may be worth while giving up time and expense in order to achieve an improved system.

2.2 Relevant Related Work on Thesauri and Ontologies

There has been recently great deal of interest and research related to thesauri. For example, the international cooperation project, AGROVOC, was undertaken by the Chinese Academy of Agricultural Information and FAO of the United Nations (Soergel et al. 2010). HILT, in United Kingdom, investigated interdisciplinary retrieval and browsing of multiple thesauri and taxonomies (Nicholson et al. 2009). In Korea a large and long-term project (2005 to 2010) aimed at producing a Korean science and technology thesaurus using conceptual and relational facets (Hanmin, Won-Kyung and Dong-In 2005). In the EU the DOPE project (Drug Ontology Project for Elsevier) aimed at exploring ways to provide access to multiple life science information sources through a single interface (Stuckenschmidt et al. 2004). Finally, in the USA, the NCI thesaurus was developed to integrate cancer research information (De Coronado et al. 2004).

This paper proposes that thesauri research should adapt to developing trends in semantic web and user needs. Because of the similarity between thesauri and ontologies, there is much ontological research that is thesauri-based. OWL and RDF-S, as significant representing methodologies, have been derived from traditional hierarchical classifications typical of thesauri, but have now become distinct from traditional thesauri conceptual understanding, as introduced by Hepp and Bruijin (2007). For instance, Ruiz-Casado et al. (2007) have put forward an automatic semantic learning model to extract semantic relations from Wikipedia.

However, the inverse type of research has seldom been attempted in the Chinese context, that is building thesauri by using ontology principles. These *onto-thesauri* can preserve the advantages of thesauri while improving a certain number of characteristics such automatic maintenance. Furthermore, onto-thesauri based on document and content management platforms would improve the recall quality and precision of information retrieval through ISO mono-lingual or multilingual indexing, homonymy resolution and the added features of ontologies (Vandamme et al. 2005).

2.3 Building Organisational Onto-thesaurus

In the context of organizations, thesauri are usually employed to develop a domain specific knowledge organization system (Mazzocchi and Tiberi 2009). The first step in building this type of knowledge organization system is to select a suitable descriptive language to construct the underlying ontology. For instance, the descriptive languages include KIF (Knowledge Interchange Format); Ontolingua; OKBC (Open Knowledge Base Connectivity); OCML (Operational Conceptual Modeling Language); Frame Logic; LOOM; and so on. With the development of Web technology, the Standard Web Languages have become the main descriptive languages, such as SHOE (Simple HTML Ontology Extension), OML (Ontology Markup Language) or XOL (XML-based ontology-exchange language). More recently several major Ontology Languages have been developed, such as CKML, OIL, DAML+OIL and OWL. In this research OWL was selected as discussed in the methodology section.

The second step consists in selecting a building method for the ontology itself. The main types of ontology building method used nowadays are as follows: text analysis, TF/IDF calculating, relation rules extracting, pattern rules matching, and RDF technology (Bagni et al. 2007, Zheng et al. 2007, Valencia-García et al. 2004). To this already extensive list some other more specific ones could be added such as the ontology construction based on a topic map-driven framework (Yang et al. 2006), or the Ontocase, which is a pattern-based ontology construction approach, or the FCA-based ontology construction for the design of class hierarchies (Blomqvist 2007, Hwang et al. 2005). There are also older approaches proposed in the 1990s that may deserve an inspection, such as the Uschold & King method, the Grüninger and Fox method and Methontology method (Lopez et al. 1999). This profusion of propositions clearly indicates that as Lopez (1999) states there is no absolute or uniform method to construct ontologies, but only suitable general approach that needs to be customized for different projects. In this research the method selected was the relation rules and map-driven framework.

The third step is focused on establishing the thesaurus relationships. These relationships include hierarchical (broader/narrower terms), equivalence (equivalent terms) and associative (related terms) relationships (Tudhope et al. 2001). This paper proposes that by merging basic concepts of ontologies with thesauri, an onto-thesaurus only requires two types of relationships, namely, hierarchical classification relationships of concepts (HC) and relationships among Thesauri Instances (RI={Y, D, S, F, C, Z}). HC refers to the generic relationships of onto-thesaurus concepts, which will be defined and discussed below as part of the onto-thesaurus construction process. The updating probability of HC will be low during the evolution process. RI refers to the relationships between standardized and non-standardized thesauri terms. These relationships include "Y," "D," "S," "F," "C" and "Z," "Y" and "D" which are defined as reverse reaction properties in the ANSI Thesauri Standard (Z39.19-2005) for Chinese corpuses. "Y" and "D" belong to equivalent relationships. "S," "F" and "Z" have the same reverse reaction properties, but they belong to hierarchical relationships, and "Z" is established by onto-thesaurus axioms. On the other hand, "C" is defined as a symmetrical property and belongs to associative relationships.

There are some commonly used methods to create the relationships of thesauri, such as the template method (Velardi et al. 2001), the conceptual clustering method (Granitzer et al. 2007), the association rules method (Nakaya et al. 2002) and mixed (template and clustering combined) method (Navigli et al. 2003). In this study, the research team adopted the template and the association rules methods as the most suitable to examine the relationships of thesauri.

Nonetheless, onto-thesauri are still thesauri in essence as the following basic comparison shows:

– Onto-thesauri are much more effective than use conceptual retrieval;
– The functions of onto-thesauri are improved as they are self-learning and auto-updating;
– Onto-thesauri have higher efficiency of updating based on the semantic relations, auto-recognition, and semi-auto maintenance.

3.0 Research methodology

This paper reports on a research project that aimed at investigating how to develop onto-thesaurus so that they have the function of auto-evolution. This is desirable in order to realize a high quality of knowledge organization and improve the recall ratio and precision ratio of information retrieval. The onto-thesaurus can then be used for information retrieval as well as a reference digital resource. So in this study, onto-thesauri encompasses two distinct layers, namely applied and technological layers, as illustrated in Figure 1. In order to develop such an onto-thesaurus, this research aimed at using an ontology model to support the construction phase and focus on neologism acquisition and semantic relationships recognition.

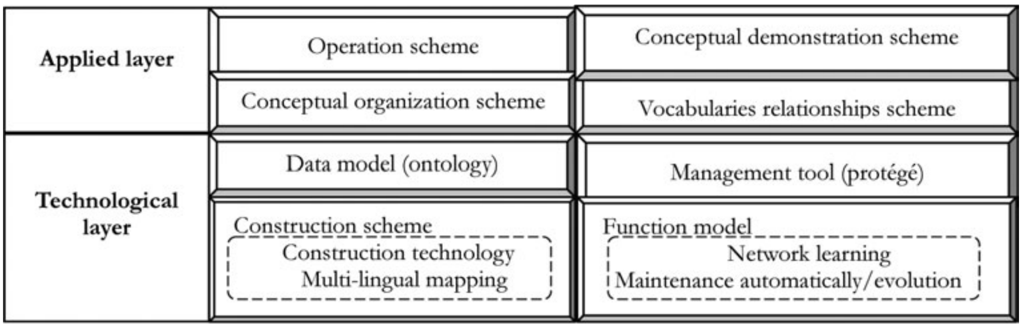


Figure 1. Logical organization of onto-thesaurus

Specifically and for the implementation, the Protégé tool and the ontology descriptive language OWL were used.

The classification used was based on *The Chinese Library Classification* (中国图书馆分类法 CLC), also known as *Classification for Chinese Libraries (CCL)*, which is effectively used as the national classification scheme in China. However, the actual experiment focused on the *Information Science and Information Service* category within the *Chinese Topic Classification Dictionary* (中国分类主题词表). Additionally, and for the purpose of neologism acquisition, the use association rule method and the template method to recognize the semantic relationships between the words, the Institute of Computing Technology Chinese Lexical Analysis System (ICTCLAS) was adopted. This is an integrated Chinese lexical analysis system that uses an approach based on multi-layer HMM. Finally, the evaluation approach was based on the Delphi method of experts' evaluation to identify best pairing of onto-thesaurus updating.

4.0 Construction and Evolution of the Chinese Information Science and Information Service (CIS&IS) Onto-Thesaurus

4.1 Definition and construction

The Chinese Information Science and Information Service (CIS&IS) Onto-thesaurus was built and evaluated based on the following definitions:

Definition 1 - Onto-thesaurus (OT) denoted by={C, I, HC, RI, O}; "OT" stands for onto-thesaurus, which contains a number of concept sets "C" (the concepts of subject category, subject classification); "I" stands for the instance subordinated to the concept sets (including term sets, standardized thesauri terms and non-standardized thesauri terms); "HC" expresses hierarchical classification relationships of concepts; "RI" symbolizes relationships among thesauri instances, and "O" represents axiom. Axiom "O" stands for identically true assertions, used for implicit relationships reasoning among thesauri instances.

Definition 2 - Hierarchical classification relationships of concepts (HC). "HC (C1, C2)" represents that "C1" is the sub-concept of "C2". For example, HC ("情报学" "information science," "情报学、情报工作" "information science and information services" means that "情报学" is the sub-concept of "情报学、情报工作." In a general case, the relationships can be expressed as "a part of."

Definition 3 - Relationships among thesauri instances (RI). "Y (I1, I2)" represents that "I1" is a non-standardized thesauri terms of "I2"; "D（I2, I1）"represents that "I2" is a standardized thesauri terms of "I1", "S (I1, I2" represents that "I1" is a lower class of "I2"; "F (I2, I1)" represents that "I2" is a upper class of "I1"; "C (I1, I2)" shows that there is an associative relationship between "I1" and "I2"; "Z（I1, I2）"represents that "I2" is a families initial term of "I1."

The 6 relationships in RI are summarized as 3 major category relationships, which are equivalent relationship (Y and D), hierarchical relationship (S, F and Z) and associative relationship (C).

The *Information Science and Information Service* category in the *Chinese Topic Classification Dictionary* has a total of 164 onto-thesaurus instances, including 8 instances in first-rate category (including 3 non-standardized thesauri terms instances); 32 instances in second-rate category; 84 instances in third-rate category; 84 instances in fourth-rate category and 6 instances in the fifth-rate category.

The definition of the CIS&IS Onto-Thesaurus is illustrated as follows:

```
<rdfs:Class
rdf:about="http://www.domain2.com#情报学与
情报工作(information science and information service) ">
```

```
<rdfs:subClassOf rdf:resource="http://www.domain2.
com#科学_科学研究(science _ scientific research)"/>
</rdfs:Class>
```

Y, D, S, F, C and Z are used to describe the relationships between thesauri conceptualizations:

```
<owl:ObjectProperty rdf:ID="D">
    <owl:inverseOf>
        <owl:ObjectProperty rdf:ID="Y"/>
    </owl:inverseOf>
    <rdfs:range rdf:resource="http://www.domain2.com#
        science _ scientific research "/>
    <rdfs:domain rdf:resource="http://www.domain2.com#
        science _ scientific research "/>
</owl:ObjectProperty>
<owl:ObjectProperty rdf:ID="Z">
    <rdfs:range rdf:resource="http://www.domain2.com#
        science _ scientific research "/>
    <rdfs:domain rdf:resource="http://www.domain2.com#
        science _ scientific research "/>
</owl:ObjectProperty>
<owl:ObjectProperty rdf:ID="S">
    <owl:inverseOf>
        <owl:ObjectProperty rdf:ID="F"/>
    </owl:inverseOf>
    <rdfs:range rdf:resource="http://www.domain2.com#
        science _ scientific research "/>
    <rdfs:domain rdf:resource="http://www.domain2.com#
        science _ scientific research "/>
</owl:ObjectProperty>
<owl:ObjectProperty rdf:about="#F">
    <owl:inverseOf rdf:resource="#S"/>
    <rdfs:range rdf:resource="http://www.domain2.com#
        science _ scientific research "/>
    <rdfs:domain rdf:resource="http://www.domain2.com#
        science _ scientific research "/>
</owl:ObjectProperty>
<owl:ObjectProperty rdf:about="#Y">
    <rdfs:domain rdf:resource="http://www.domain2.com#
        science _ scientific research "/>
    <rdfs:range rdf:resource="http://www.domain2.com#
        science _ scientific research "/>
    <owl:inverseOf rdf:resource="#D"/>
</owl:ObjectProperty>
<owl:DatatypeProperty rdf:about="http://www.domain2.
    com#hasChineseTerm">
    <rdfs:range
        rdf:resource="http://www.w3.org/2001/XMLSchem
        a#string"/>
    <rdfs:domain rdf:resource="http://www.domain2.com#
        science _ scientific research "/>
</owl:DatatypeProperty>
<owl:SymmetricProperty rdf:ID="C">
    <owl:inverseOf rdf:resource="#C"/>
    <rdf:type rdf:resource="http://www.w3.org/2002/07/
        owl#ObjectProperty"/>
    <rdfs:domain rdf:resource="http://www.domain2.com#
        science _ scientific research "/>
    <rdfs:range rdf:resource="http://www.domain2.com#
        science _ scientific research "/>
</owl:SymmetricProperty>
```

A term which is subordinate to a certain thesaurus instance can be presented by using a concept instance, expressed as follows:

```
<thes: 情报检索 (information retrieval) rdf:about="http://www.
domain2.com# 查全率（recall ratio） ">
    <thes:hasChineseTerm
    rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >查全率（recall ratio） </thes:hasChineseTerm>
    </thes: 情报检索 (information retrieval)>
```

In terms of illustration, if 查全率（recall ratio）is a thesaurus term of 情报检索 (information retrieval), thes:has ChineseTerm is a value attribute of 查全率（recall ratio), meanwhile 查全率（recall ratio）is an attribute value of information retrieval. The advantage of this approach is that the more value attributes added, the better for thesauri description. Furthermore, this approach would also support a truly multilingual thesaurus by adding a value attribute <thes:hasEnglishTerm> as a guiding role.

4.2 Data source

4.2.1 Data source acquisition

The corpus used for the empirical research was the *Chinese Knowledge Infrastructure Database* (CNKI), which includes 19 specific academic journals on the subject of library and information science. This data source consists of more than 400 documents. An illustration is shown in Figure 3, including the document's title, keywords abstract and navigator.

In this study, a traversal algorithm is designed to describe the information of documents using the parameters of filename and database name (dbname), as follows:

"database name + database publish year" and "name of Journal + Journal publish year + issue number + volume number and/or year".

For example,

http://dlib.cnki.net/kns50/detail.aspx?dbname=CJFD2013&filename=TSGT20130507

id	title	keywords	abstract	navigator
277	2000▮2007年知识网络研究分...	知识网络;文献研究;文献计量...	以2000▮2007年ISI Web of Science数据库收录的知识网络研究文献为研究对...	文化、科学、教育、体育
278	20年可视化发展历程对情报...	可视化;信息组织;信息检索;...	可视化是情报学研究中涌现出的一个新热点.本文通过对可视化发展20年历...	文化、科学、教育、体育
279	21世纪情报学的发展与应用...	情报学;学科建设;发展研究;	随着社会信息化进程的加速推进,情报学的重要性日益增加.情报服务与情...	文化、科学、教育、体育
280	Internet科技信息资源门户网...	STIP;搜索引擎;多线程搜索;...	Internet是全球最大而且还在迅速增长的信息资源库，有效地组织和开发¶¶...	文化、科学、教育、体育
281	20年可视化发展历程对情报...	可视化;信息组织;信息检索;...	可视化是情报学研究中涌现出的一个新热点.本文通过对可视化发展20年历...	文化、科学、教育、体育
282	Internet科技信息资源门户网...	STIP;搜索引擎;多线程搜索;...	Internet是全球最大而且还在迅速增长的信息资源库，有效地组织和开发¶¶...	文化、科学、教育、体育
283	Starlogo对Wiki演化过程模拟...	Starlogo;Wiki;演化;模拟;	本文以Wikipedia为例,概括了Wiki演化过程的主要特点.认为用户数量的增长...	文化、科学、教育、体育
284	Web2.0有序性的应用▮Lib2.0	Web2.0;Lib2.0;自组织;有序性;	本文主要探讨Web2.0的自组织和有序性在图书馆中的应用所形成的Lib2.0...	文化、科学、教育、体育
285	Web信息服务中受控语言研究	受控语言;信息服务;信息检索...	Web信息服务采用知识组织体系和规范操作流程构建基于主题的有序信息...	文化、科学、教育、体育
286	WWW科技信息资源自动标引...	自动分类;自动标引;自动分词...	随着因特网的迅速发展和广泛应用，它逐渐成为目前最大的信息资源宝库...	文化、科学、教育、体育
287	一种网络信息资源综合评价...	网络资源;信息资源;评价指标...	本文在构建网络信息资源评价指标体系的基础上,提出了一种基于层次分析...	文化、科学、教育、体育
288	上海情报学硕士教育发展现...	情报学;研究生教育;上海;	对上海8个情报学硕士招生单位进行调研.主要内容包括培养目标、研究方...	文化、科学、教育、体育
289	不正当竞争情报行为及其防...	竞争情报;不正当竞争情报行...	在▮知讯者生存▮的信息时代，企业的情报意识和情报能力日益增强，竞争...	经济 经济计划与管理
290	个性化推荐系统中基于本体...	用户建模;个性化;本体;可扩...	随着信息的急剧膨胀，▮信息过载▮、▮资源迷向▮等问题相继出现，信息检索...	文化、科学、教育、体育
291	个性化知识服务中基于Ontolo...	Ontology;用户兴趣模型;个性...	个性化知识服务是知识经济时代信息服务发展的必然趋势.是满足信息用户...	文化、科学、教育、体育
292	中国情报学研究的继承与发...	情报学;继承;发展;中国;	本研究以维普期刊论文数据库为数据来源.利用专用软件分析1990年至2005...	文化、科学、教育、体育
293	中日情报学专业研究生课程...	情报学;课程体系;教育模式;...	本文针对情报学专业研究生教育的状况,把中日情报学专业研究生课程体系...	
294	中美图书情报用户教育发展...	图书情报用户教育;文献资源;...	通过中美图书情报用户教育发展概况、形式和内容.以及两国的教学规模进...	文化、科学、教育、体育
295	中美大学网络影响因子研究	网络影响因子;大学网站;回归...	选取了中国和美国26所大学,对其进行了网络影响因子测定,考察了其中6项...	文化、科学、教育、体育
296	人际网络开发与竞争情报发展	人际网络;竞争情报;信息转化...	人际网络开发事关情报和竞争情报发展的大局。从信息转化理论和两种信...	文化、科学、教育、体育
297	从情报学与相关学科的差异...	情报学;情报技术;相关学科;	情报学与图书馆学、信息管理、通信科学、智能科学、计算机科学、计算...	文化、科学、教育、体育
298	企业竞争情报与知识管理的...	竞争情报;知识管理;系统整合;	随着经济全球化的发展和知识经济时代的来临，竞争情报和知识管理迅速...	经济 经济计划与管理
299	信息化时代医务人员的情报...	医务人员;信息素养;信息社会;	阐明了情报信息素养的概念.指出了信息素养是信息化社会对医生提出的新...	医药、卫生 预防医学、卫
300	信息可视化系统通用框架设计	信息可视化;系统框架;通用系...	对信息可视化系统通用框架问题进行了研究和探讨.介绍并分析了系统框架...	文化、科学、教育、体育
301	信息哲学与图书情报学	信息;信息哲学;图书情报学;	从介绍弗罗里迪的信息哲学和图书情报学的关系入手.围绕信息哲学对图书...	文化、科学、教育、体育
302	信息时代对情报学的影响	信息时代;情报学;学科发展;	随着信息技术的不断进步.信息时代对情报学的影响也越来越大.文章对比...	文化、科学、教育、体育
303	信息构建在情报学领域的应...	信息构建;情报学;	本文对信息构建的核心内容、特点进行了概述.分析了情报学研究现状.然...	文化、科学、教育、体育
304	信息熵在竞争情报计量分析...	信息熵;竞争情报;主题研究;...	采用信息熵分析方法.以情报科学《十年来我国竞争情报文献的计量分析》...	文化、科学、教育、体育

Figure 2. Data source.

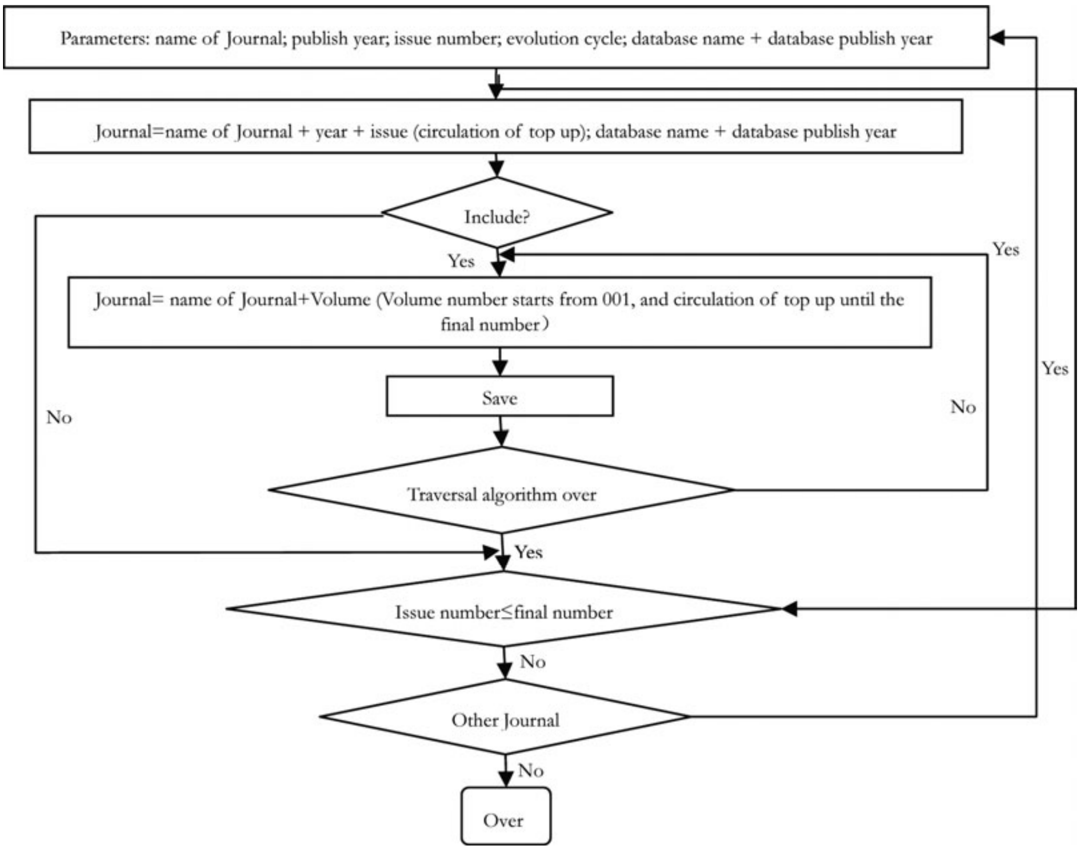


Figure 3. The traversal algorithm.

is the descriptive information about the journal *Library*, Issue 5, published in 2013, 7th article. Therefore a regulation was established that the first half of URL is fixed whilst the last half (dbname and filename) are specific to the article processed, further explanation is provided in Figure 2.

The data source is crucial for the neologisms acquisition and semantic relationships recognition. Neologisms derive mainly from frequencies of occurrence in titles, keywords and abstracts whilst the semantic relationships are mainly drawn from syntactic construction in the abstract.

4.2.2 Words processing and neologisms acquisition

In this study, document (that is, journal article) refers to title, abstract and keywords analyzed during the word and neologisms acquisition process. The first step to achieve this acquisition, ICTCLAS starts by analyzing the segment "Keywords" in the document. After this, if the title of the document contains some of the already processed keywords then these should be eliminated in order to reduce search and error frequency. In addition, the title should be segmented as follow: the segmentation results of the title "叙词本体构建及其演化模型研究 (research on construction and evolution of onto-thesaurus)" are as: 叙/j 词/n 本体/n 构建/v 及其/c 演化/vn 模型/n 研究/vn. In this way the result of word segmentation and part of Chinese specific speech tagging is achieved.

As part of the second step, noun/n, name form word/an, object form/Ng and nominal verb/vn are extracted from the segmentation results. The third step, eliminates general words, references the general wordlist, and then conducts frequency statistics. Finally, the new words are added to the general wordlist, frequency statistics are saved into the database as well as the issue number of origin article.

As stated above, in this research the standard of neologisms acquisition is based on the *Chinese Topic Classification Dictionary*. This dictionary's regulation defines that if the word times appeared in a document is up to 5 times in titles or 10 times in keywords or 15 times in abstracts, it can be used as a formal topic word. According to this regulation, the title is set as a benchmark, a neologism reference standard is recognized if it appears 5 times in titles or 10 times in keywords or 15 times in abstracts.

4.3 Semantic relationships recognition

The process to identify relationships of neologisms consists of two steps. Firstly, the association rules method is applied for obtaining associative relationships. Secondly, template method is used to identify hierarchical relationships and equivalence relationships.

When calculating frequency of the word, different weight coefficients were devised according to the different positions where words appeared in the document. So that weight coefficient of title is 1.0, weight coefficient of keywords is 0.5, weight coefficient of abstract is 0.3. For instance, if one word appears 1 time in a document title, 1 time in the keywords, 2 times in abstract, the value of the final word frequency is calculated as $1+0.5+0.3\times2$, which equals 2.1. So in this research, according to the neologism reference standard, the threshold value of the word frequency is defined as 5.0.

4.3.1 Associative relationships

Firstly, if two terms appear in the same chapter fragment (such as full-text, chapter, section, paragraphs or sentence), then a relationships between the neologism and the other words in the abstract appear in the same document can be identified through the algorithm of the association rules method from the calculating co-occurrence frequency. Furthermore, if the co-occurrence frequency of terms is higher than the threshold value, these two terms are to be preliminarily recognized as an associative relationship (Maedche and Staab 2001).

The definition of association rules method can be described as a document sets $T = \{ti \mid i=1... n\}$, each document contains a series of terms $ti= \{j \mid j = 1... m, j\in$ conceptual sets $C\}$. If the co-occurrence frequency of X_k and Y_k (X_k, $Y_k \in C$, $X_k \cap Y_k$= null) in document ti (i= 1... n) is higher than the threshold value, it can be proved that X_k and Y_k have some related relationship.

The support degree of association rules is defined as a percentage of a subset which appeared co-occurrence in $X_k U Y_k$ in a document. The confidence degree of association rules refers to a percentage of Y_k appeared in the document when X_k appeared. If the support and confidence degree of association rules are higher than the threshold value, it means X_k and Y_k have a associative relationship.

4.3.2 Hierarchical and equivalence relationships

The template method means analysis on statistics and machine learning for documents in a certain domain (Daille 1996). Firstly, conceptual relationships are constructed according to syntax rules. Secondly, word-sequence is extracted from the corpus. In this study, the definition of template method adopts the syntax form of regular expression. 16 syntax rules were defined for this study, characterizing 6 equivalence relationships and 10 hierarchical relationships.

An equivalence relationship is a substitution relationship among thesauri which has the same or similar semantic, also known as identical relationship or substitutive relationship (Ma 2003). For example, the term "I1" is used as another non-specification form to term "I2;" in this case, "Y(I1,I2)" means that "I2" is the standard expression with "I1," and the relationship "D（I2,I1)" means that "I1" is the substitutive expression with "I2" (with its non-specification form).

Regulation 1：WORD0亦称 {WORD1, WORD2… (and|or)}WORDn
Then D (WORD0, WORDi), Y (WORDi, WORD0), (i∈ {1，2，…n});

Regulation 2：WORD0 {又|俗|也|或} 称 {WORD1, WORD2… (and|or)}WORDn
Then D (WORD0, WORDi) ，Y (WORDi, WORD0), (i∈{1，2，…n});

Regulation 3：WORD0 {又|亦|也|或} 叫 {WORD1, WORD2... (and|or)}WORDn
Then D (WORD0, WORDi), Y (WORDi, WORD0), (i∈ {1，2，…n});

Regulation 4：WORD0是 {WORD1, WORD2… (and|or)}WORDn的另一种 {规范|严格}{表达方式|称谓|称呼}
Then D (WORD0, WORDi), Y (WORDi, WORD0), (i∈ {1，2，…n});

Regulation 5：｛WORD1, WORD2… (and|or)}WORDn 是WORD0的 {另一种|其他} {表达方式|称谓|称}
Then D (WORDi, WORD0), Y (WORD0, WORDi), (i∈ {1，2，…n});

Regulation 6：WORD0, 即 {|是} {WORD1, WORD2 … (and|or)} WORDn
则D(WORD0，WORDi), Y(WORDi,WORD0) ,(i∈{1，2，…n})

Hierarchical relationships are established between lower class and upper class, also known as S and F (Ma 2003). If term "I1" can be used as one type, a part or a body to term "I2," then it can be defined as S (I1, I2) and F（I2, I1).

Regulation1：WORD0, 比如 (for example) {WORD1, WORD2… (and|or)}WORDn
Then S (WORDi，WORD0), (WORD0, WORDi);

Regulation2：像 {WORD1, WORD2… (and|or)}WORDn 这样的WORD0
Then S (WORDi, WORD0), F (WORD0, WORDi), (i∈ {1，2，…n});

Regulation3：｛WORD1, WORD2… (and|or)}WORDn,{或者|以及|或|和|及}其他的WORD0
Then S (WORDi, WORD0), F (WORD0, WORDi), (i∈ {1，2，…n});

Regulation4：WORD0, {包括|包含} {WORD1, WORD2… (and|or)}WORDn
Then S (WORDi, WORD0), F (WORD0, WORDi), (i∈ {1，2，…n});

Regulation5：WORD0, 尤其是 {WORD1, WORD2… (and|or)}WORDn
Then S (WORDi, WORD0), F (WORD0, WORDi), (i∈ {1，2，…n});

Regulation 6：｛WORD1, WORD2… (and|or)}WORDn 是 {一种} WORD0*
Then S (WORDi, WORD0), F (WORD0, WORDi), (i∈ {1，2，…n});

*Regulation 7：｛WORD1} {是指|指的是|指} *WORD0*
Then S (WORD1, WORD0), F (WORD0, WORD1);

Regulation 8：WORD0 {分为|划分为|分割成|分成|划分成} {WORD1, WORD2... (and|or)} WORD0
Then S (WORDi, WORD0), F (WORD0, WORDi), (i∈ {1，2，…n});

Regulation 9：｛WORD1, WORD2… (and|or)}WORDn {属于|隶属于} WORD0
Then S (WORDi, WORD0), F (WORD0, WORDi), (i∈ {1，2，…n});

Regulation 10：WORD0 有 {WORD1, WORD2… (and|or)}WORDn 几种类型
则S(WORDi,WORD0),F(WORD0,WORDi) , (i∈{1，2，…n})

4.4 Evolution Mechanism

4.4.1 Expert Evaluation

Delphi Method of expert evaluation mechanism (Linstone and Turoff 2002) was adopted to improve the evolution function of onto-thesaurus being developed by this study. In this research, the experts are in the field of library and information science. And the e-mail is used as the main review way to send message to the experts. The experts review is as follows: Firstly, the mechanism can be defined by using scoring from fifteen experts to identify the neologisms and the relationships between them. Secondly, two authoritative experts have the final voting rights and are entitled to exercise a veto with only one vote on all different opinions from other experts. Thirdly, the results can be measured from the arithmetic mean of the optional schemes, as illustrated in Table 1.

Experts	Optional schemes					
	1	2	…	j	…	n
1	C_{11}	C_{12}	…	C_{1j}	…	C_{1n}
2	C_{21}	C_{22}	…	C_{2j}	…	C_{2n}
…	…	…	…	…	…	…
I	C_{i1}	C_{i2}	…	C_{ij}	…	C_{in}
…	…	…	…	…	…	…
M	C_{m1}	C_{m2}	…	C_{mj}	…	C_{mn}

Table 1. Score chart

From the Table 1 the arithmetic mean value of all schemes can be calculated according to:

$$M_j = 1/m_j \sum C_{ij} \quad (1)$$

M_j is the arithmetic mean value of j scheme (j=1, 2… n). m_j stands for the number of experts who participated in f scheme. C_{ij} stands for the value that i expert set for j scheme (i=1,2,……,m). From the result, the scheme which has the maximum arithmetic mean value can be chosen as the recommended scheme.

4.4.2 Updating the ontology database with expert feedback

The neologisms and semantic relationships need to be updated into ontology database after the evolution mechanism. This project used the Jena package for the implementation of the application and the database. Jena provides a method to define class, attributes and instances on ontologies. The method is described as follows:

```
//First, create the object myModel of Model.
    Model myModel= ModelFac-
        tory.createDefaultModel();
//Applying "read" method in class Model to realize
    query of onto-thesaurus.
    FileInputStream file=new FileInput-
        Stream("E:\\thesauri.owl");
    InputStreamReader in=new InputStream-
        Reader(file,"UTF-8");
    myModel.read(in, null);
    in.close();
```

Neologisms and semantic relationships stand for instances and attributes respectively in onto-thesauri. The method of adding is as follows:

```
//Adding neologisms.
    Resource rs=myModel.createResource
        ("http://www.scucl.com#情报学 (information sci-
        ence)");
     Individual firstInd=myModel.createIndividual
     ("http://www.scucl.com#情报观(information sci-
     ence view), rs);
    modelAction.writeModel(myModel);
//Adding semantic relationships.
     Individual first-
     Ind=myModel.getIndividual("http://www.scucl.com
     #情报观view (information science)");
     Individual second-
     Ind=myModel.getIndividual("http://www.scucl.com
     #信息观 (information view)");
    Property prop=myModel.getProperty("C");
    firstInd.addProperty (prop, secondInd);
    modelAction.writeModel (myModel);
```

5.0 Results, analysis and discussion

5.1 Processing words

The evaluation indicators of entity extraction include precision (P), the rate of recall (R) and F-score value. The definitions of these indicators are as follows:

$$P = \frac{N_1}{N_1 + N_2} \quad (2)$$

$$R = \frac{N_1}{N_1 + N_3} \quad (3)$$

N_1 means that the number of words correctly identified. N_2 expresses the number of words incorrectly identified as words of other properties that are mapped to this property. N_3 is the number of words which belongs to one property but is mistaken as others.

$$F = \frac{(\beta^2 + 1) \times P \times R}{\beta^2 \times P + R} \quad (4)$$

Parameter P is the weight of P and R. In this experiment, β=1, which indicates the same weight to P and R. The number of non-general words reached 530 from the experiment, and the amount of general words is summarized as 333, of which 31 are under-standard. So P is calculated as P = (530 – 31)/530 = 94.15%.

If the terminology appear in the title and abstract simultaneously, we hold the keywords as a terminology, not as a neologism. So the word frequency should be revised and segmented after selecting the terminology out from title. After manual checking and experts review, only 30 thesauri terms are not recognized in the experiment system. So the R is calculated as R = (530 – 31)/(530 – 31) + 30 = 94.33% and F = (1+1) × 94.15% × 94.33%/ (94.15% + 94.33%) = 94.24%. It seems that the test result is good.

In this study, words which are defined as widely used if the rate of word frequency is 5.0 or higher. But there are only 22 words matching this defined regular within 500 words, and almost about 20 words are neologisms through comparing the onto-thesaurus database.

5.2 Semantic relationships

If the terms appear simultaneously in the abstracts of multiple documents, and the co-occurrence has more frequency than the threshold value, then they can be regarded as have some related relationship. In this experiment, the co-occurrence frequency threshold is set as 5.0, a total of 22 pairs of the relationship recognition are chosen, and 12 groups of words appeared in 2 documents and, 7 groups of words appeared in 3 documents, 2 groups appeared in 4 documents and 1 group appeared in 5 documents. Through a further template method matching experiment, only 2 groups are identified as "D." The results can be seen in Figure 4.

5.3 Experiment system results

The thesaurus of this experiment is based on the basic and regular sources of CIS&IS. So the accuracy of this system can be verified from the manual analysis and experts view based on the Chinese thesaurus regulations.

> 网络，评价：(network, evaluation) 在网络 (network) 信息资源的评价 (evaluation) 中，专家给出的信息可能为纯语言信息。

Id	subject items	related items	ecids	relationship types	co-occurrence frequency
100	信息检索(information retrieval)	情报学(information science)	278,281,376,413,447	参(C)	5
101	信息检索(information retrieval)	信息(information)	280,282	参(C)	2
102	信息检索(information retrieval)	资源(resource)	280,282	参(C)	2
104	信息(information)	信息检索(information retrieval)	280,282	参(C)	2
105	信息(information)	资源(resource)	280,282,286	参(C)	3
108	信息(information)	情报(intelligence)	363,370,378	参(C)	3
112	资源(resource)	信息检索(information retrieval)	280,282	参(C)	2
113	资源(resource)	信息(information)	280,282,286	参(C)	3
118	网络(network)	评价(evaluation)	351,443	代(D)	2
120	竞争情报(Competitive Intelligence)	知识管理(knowledge management)	298,348,427	参(C)	3
129	发展(development)	情报学(information science)	292,373,375	代(D)	3
130	知识管理(knowledge management)	竞争情报(Competitive Intelligence)	298,348,427	参(C)	3
131	知识管理(knowledge management)	情报学(information science)	311,413,437,457	参(C)	4
139	情报(intelligence)	图书馆(library)	330,363	参(C)	2
140	情报(intelligence)	信息(information)	363,370,378	参(C)	3
141	情报(intelligence)	情报学(information science)	370,379	参(C)	2
144	领域(domain)	情报学(information science)	303,343,367	参(C)	3
148	评价(evaluation)	网络(network)	351,443	参(C)	2
159	指标体系(indicator system)	期刊评价(Journal evaluation)	324,395	参(C)	2
164	图书馆(library)	情报(intelligence)	330,363	参(C)	2
167	期刊评价(Journal evaluation)	指标体系(indicator system)	324,395	参(C)	2
173	网络环境(network environment)	情报学(information science)	370,437	参(C)	2

Figure 4. Results of Semantic relationships recognition

> 发展，情报学: (development, information) 论述了情报学 (information science) 体系结构和未来的演化及发展趋势, 并从哲学的角度分析了情报学的发展 (development) 规律。

Actually, apart from the others which are verified correctly from the experiment system, there are 2 groups recognized as the "D" relationships (网络and 评价 ("network" and "evaluation"), 发展and 情报学 ("development" and "information science")) , which are obviously not have the relationship from the Chinese thesauri regulations. So the accuracy of the evolution system should be researched for a further step.

6.0 Conclusion

In this study, under the guidance of the related theory of thesauri and ontology, an exploratory study was conducted to analyze "onto-thesaurus," and CIS&IS evolution system is the first performed research.

- First, the conceptual relationships of onto-thesauri can be extended to cover the shortage of the traditional 6 thesaural relationships.
- Second, based on the corpus, the automatic maintenance function would be achieved.
- Third, onto-thesaurus can be used as information retrieval corpus and internet search engines, and this would lead to an increase in the retrieval efficiency in knowledge organization.

Experiments indicated that, by using natural language processing technology, calculating frequency and segmenting words, our method achieved good results, with a rate of recall of 94.33% and *F*-score values of 94.24% in the neologism collections. Moreover, to determine the relationships among onto-thesauri, association rules method and the template method were used to recognize the associative relationship, equivalence relationship and hierarchical relationship. Meanwhile, an expert review method was presented to reduce the inaccurate probability of the experiment.

So the experimental results indicated that the method proposed in this paper is more suitable for improving the functionality of thesauri based on ontological features.

Some problems remain to be solved in future work. Only 16 regular expressions were defined, which cannot cover all grammatical rules. In addition, if the statement of the sentence is not standardized with the templates, some relationships can easily be missed. So the success rate of identifying hierarchical relationships and equivalence relationships based on a template method did not reach the ideal state. Furthermore, during the semantic relationships recognition process, the relationships between neologisms and thesauri in the onto-thesaurus database were not considered already, only relationships among the neologisms have been accomplished. From the above reasons, it is worthwhile trying to further analyze word segmentation, knowledge acquisition and syntactic parsing, to realize the onto-thesaurus evolution system much more completely.

References

ANSI-NISO Z39.19-2005. 2005. *Guidelines for the construction, format, and management of monolingual controlled vocabularies*.

Antoniou Grigoris and Van Harmelen, Frank. 2004. *A semantic web primer*. Cambridge, MA: MIT Press.

Antoniou Grigoris and Van Harmelen Frank. 2009. *Web ontology language: OWL. Handbook on ontologies*. Berlin, Heidelberg: Springer.

Bagni, Daniele, Cappella, Marco, Pazienza, Maria Teresa, Pennacchiotti, Marco and Stellato, Armando. 2007. Harvesting relational and structured knowledge for ontology building in the WPro architecture. In Basili, Roberto and Pazienza, Maria Teresa, eds., *Proceedings of AI*IA 2007: artificial intelligence and human-oriented computing: 10th Congress of the Italian Association for Artificial Intelligence, Rome, Italy, September 10-13, 2007,* Berlin; New York: Springer, pp. 157-69.

Bechhofer, Sean, Van Harmelen, Frank, Hendler, Jim, Horrocks, Ian, McGuinness, Deborah L., Patel-Schneider, Peter F. and Stein, Lynn Andrea. 2004. *OWL web ontology language reference W3C recommendation 10 February 2004*. Available http://www.w3.org/TR/owl-ref/

Blomqvist, Eva. 2007. OntoCase-a pattern-based ontology construction approach. In *Proceedings of OTM: ODBASE-The 6th International Conference on Ontologies, DataBases, and Applications of Semantics, Vilamoura, Portugal,* 4803: 971-88.

Daille, Beatrice. 1996. *Study and implementation of combined techniques for automatic extraction of terminology. The balancing act: combining symbolic and statistical approaches to language*. Cambridge, MA: MIT Press.

De Coronado, Sherri, Haber, Margaret W., Sioutos, Nicholas, Tuttle, Mark S. and Wright, Lawrence W. 2004. NCI Thesaurus: using science-based terminology to integrate cancer research results. *Medinfo*, 11: 33-7.

4th European Semantic Web Conference, ESWC 2007, Innsbruck, Austria, June 3-7, 2007. Proceedings. Lecture notes in computer science 4519, Heidelberg, Berlin: Springer, pp. 129-44.

Gilchrist, Alan. 2003. Thesauri, taxonomies and ontologies – an etymological note. *Journal of documentation* 59: 7-18.

Granitzer, Michael, Scharl, Arno, Weichselbraun, Albert, Neidhart, Thomas, Juffinger, Andreas and Wohlgenannt, Gerhard. 2007. Automated ontology learning and validation using hypothesis testing. In Wegrzyn-Wolska, Katarzyna M and Szczepaniak, Piotr S., eds., *Advances in intelligent web mastering: proceedings of the 5th Atlantic Web Intelligence Conference-AWIC'2007, Fontainebleau, France, June 25-27, 2007*, Berlin; New York: Springer, pp. 130-5.

Gruber, Thomas R. 1993. A translation approach to portable ontology specifications. *Knowledge acquisition* 5: 199-220.

Gruber, Thomas R. 2009. Ontology. In Özsu, M Tamer and Liu, Ling, eds., *Encyclopedia of database systems*, New York; London: Springer.

Hepp, Martin and Bruijin, Jos de. 2007. GenTax: a generic methodology for deriving OWL and RDF-S ontologies from hierarchical classifications, thesauri, and inconsistent taxonomies. In Franconi, Enrico, Kifer, Michael and May, Wolfgang, eds., *The Semantic Web: Research and Applications*

Hitzler, Pascal, Krötzsch, Markus, Parsia, Bijan, Patel-Schneider, Peter F., Rudolph, Sebastian. 2009. *OWL 2 web ontology language primer W3C recommendation 27 October 2009*. Available http://www.w3.org/2009/pdf/REC-owl2-primer-20091027.pdf.

Horrocks, Ian and Patel-Schneider, Perter F. 2004. Reducing OWL entailment to description logic satisfiability. *Web semantics: science, services and agents on the World Wide Web* 1: 345–57.

Hwang, Suk-Hyung, Kim, Hong Gee and Yang, Hae Sool. 2005. A FCA-based ontology construction for the design of class hierarchy. Gervasi, Osvaldo, Gavrilova, Marina L., Kumar, Vipin, Laganà, Antonio, Lee, Heow Pueh, Mun, Youngsong, Taniar, David and Tan, Chih Jeng Kenneth, eds., *Computational Science and Its Applications – ICCSA 2005 International Conference, Singapore, May 9-12, 2005, Proceedings, Part III.* Lecture notes in computer science 3482, Heidelberg, Berlin: Springer, pp. 827-35.

Hanmin, Jung, Won-Kyung, Sung and Dong-In, Park. 2005. Project report on a Korean science & technology thesaurus with conceptual/relational facets. In Sojka, Petr, Choi, Key-Sun, Fellbaum, Christiane and Vossen, Piek, eds., *GWC 2006, Proceedings*, pp. 311-3.

Kalinichenko Leonid, Missikoff Michele, Schiappelli Federica and Skvortsov Nikolay. 2003. Ontological modeling. In *Proceedings of the 5th Russian Conference on Digital Libraries (RCDL 2003), Saint-Petersburg, Russia, 29 – 31 October 2003*

Kless, Daniel, Milton, Simon and Kazmierczak, Edmund. 2012. Relationships and relata in ontologies and thesauri: differences and similarities. *Applied Ontology* 7: 401-28.

Linstone, Harold A. and Turoff, Murray. 2002. *The Delphi method techniques and applications reading*. MA: Addison-Wesley.

Lopez, Mariano Fernandez. 1999. Overview of methodologies for building ontologies. In Benjamins, V Richard, ed., *Proceedings of the IJCAI-99 workshop on Ontologies and problem-solving methods: lessons learned and future trends.* Amsterdam: Aachen: University of Amsterdam.

Lopez, Mariano Fernandez, Gomez-Perez, Asuncion and Sierra, Juan Pazos. 1999. Building a chemical ontology using methontology and the ontology design environment. *Intelligent systems and their applications* 14 n.1: 37-46.

Losee, Robert M., 2007. Decisions in thesaurus construction and use. *Information processing and management* 43: 958-68.

Ma, Zhanghua. 2003. *Information organization*. Beijing: Tsinghua University Press.

Maedche, Alexander and Staab, Steffen. 2001. Ontology learning for the semantic web. *IEEE intelligent systems* 16 n.2: 72-9.

Marcondes, Carlos Henrique. 2013. Knowledge organization and representation in digital environments: relations between ontology and knowledge organization. *Knowledge organization* 40: 115-22.

Mazzocchi, Fulvio and Tiberi Melissa. 2009. Knowledge organization in the philosophical domain: dealing with polysemy in thesaurus building. *Knowledge organization* 36: 103-12.

Nakaya, Naomi, Kurematsu, Maski and Yamaguchi, Takahira. 2002. A domain ontology development environment using a MRD and text corpus. In Welzer, Tatjana, Yamamoto, Shuichiro and Rozman, Ivan, eds., *Knowledge-based software engineering: proceedings of the fifth Joint Conference on Knowledge-Based Software Engineering,* Amsterdam; Washington, DC: IOS Press/Ohmsha, pp. 242-54.

Navigli, Roberto, Velardi, Paola and Gangemi, Aldo. 2003. Ontology learning and its application to automated terminology translation. *IEEE intelligent systems* 18 n.1: 22-31.

Nicholson, Dennis, McCulloch, Emma and Joseph, Anu. 2009. *HILT: high-level thesaurus project. phase IV and embedding project extension: final report.* Available http://eprints.rclis.org/13186/1/strathprints014046.pdf.

Panzer, Michael and Zeng, Marcia Lei. 2009. Modeling classification systems in SKOS: some challenges and best-practice recommendations. In DC-2009 International Conference on Dublin Core and Metadata Applications, 12-16 October 2009, the National Library of Korea, Seoul: "semantic interoperability of linked data." Seoul: Korean Library Association, pp. 3-14.

Pastor-Sanchez, Juan-Antonio, Mendez, Francisco Javier Martinez and Rodriguez-Munoz, Jose Vicente. 2009. Advantages of thesauri representation using the simple knowledge organization system (SKOS) compared with proposed alternatives. *Information research* 14. Available http://www.informationr.net/ir/14-4/paper422.html.

Rosati, Riccardo. 2005. Semantic and computational advantages of the safe integration of ontologies and rules. In Fages, François and Soliman, Sylvain, eds., *Principles and Practice of Semantic Web Reasoning Third International Workshop, PPSWR 2005, Dagstuhl Castle, Germany, September 11-16, 2005 Proceedings*. Lecture notes in computer science 3703. Berlin Heidelberg: Springer: 50-64.

Ruiz-Casado, Maria, Alfonseca, Enrique and Castells, Pablo. 2007. Automatising the learning of lexical patterns: An application to the enrichment of WordNet by extracting semantic relationships from Wikipedia. *Journal of data and knowledge engineering* 61: 484-99.

Soergel, Dagobert, Lauser, Boris, Liang, Anita, Fisseha, Frehiwot, Kerzer, Johannes and Katz, Stephen. 2010. Reengineering thesauri for new applications: the AGRO VOC example. *Journal of digital information* 4 n.4: 1-23.

Stuckenschmidt, Heiner, van Harmelen, Frank, De Waard, Anita, Scerri, Tony, Ravinder, Bhogal, van Buel, Jan, Crowlesmith, Ian, Fluit, Christiaan, Kampman, Arjohn, Broekstra, Jeen and van Mulligen, Erik. 2004. Exploring large document repositories with RDF technology: the DOPE project. *IEEE intelligent systems* 19 n.3: 34-40.

Sun, Qian and Wan, Jiancheng. 2007. Study on approach of building domain ontology based on thesaurus. *Computer engineering and design* 28: 5054-6.

Tang, Aimin, Zhen, Zhen and Fan, Jing. 2005. Thesaurus based approach to build domain ontology. *New technology of library and information service* 4: 1-5.

Tang, Jing. 2004. Research on transforming thesauri into ontology. *Information studies: theory and application* 27: 642-5.

Tudhope, Douglas, Alani, Haith and Jones, Christopher. 2001. Augmenting thesaurus relationships: possibilities for retrieval. *Journal of digital information* 1 n.8. Available http://eprints.soton.ac.uk/254484/1/Tudhope_JoDI.pdf

Valencia-García, Rafael, Castellanos Nieves, Dagoberto, Vivancos Vicente, Pedro José, Fernández Breis, Jesualdo Tomás, Martínez-Béjar, Rodrigo and García Sánchez, Francisco. 2004. An Approach for ontology building from text supported by NLP techniques. In Conejo, Ricardo, Urretavizcaya, Maite and Pérez-de-la-Cruz, José-Luis, eds., *Current Topics in Artificial Intelligence 10th Conference of the Spanish Association for Artificial Intelligence, CA-EPIA 2003, and 5th Conference on Technology Transfer, TTIA 2003, San Sebastian, Spain, November 12-14, 2003. Revised Selected Papers*. Lecture notes in computer science 3040. Berlin Heidelberg: Springer, pp. 126-35.

Vandamme, F., Wang, L., Vandamme, M. and Kaczmarski P. 2005 Onto-thesauri: an efficient ontology! In Li, Daoliang, Wang, Baoji, eds., *Artificial Intelligence Applications and Innovations IFIP TC12 WG12.5 - Second IFIP Conference on Artificial Intelligence Applications and Innovations (AIAI2005), September 7–9, 2005, Beijing, China*, Springer, pp.41-47.

Velardi, Paola, Fabriani, Paolo and Missikoff Michele. 2001. Using text processing techniques to automatically enrich a domain ontology. In *Proceeding FOIS '01 Proceedings of the International Conference on Formal Ontology in Information Systems*, New York: ACM, pp. 270-84.

White, Hollie, Willis, Craig and Greenberg, Jane. 2013. HIVEing: the effect of a semantic web technology on inter-indexer consistency. *Journal of documentation* 70: 1-43.

Willis, Craig and Losee, Robert M. 2013. A random walk on an ontology: using thesauri structure for automatic subject indexing. *Journal of the American Society for Information Science and Technology* 64: 1332-44.

Yang, Cheng-Zen, Chen, Ing-Xiang, Chou, Chun-Hua and Yang, Meng-Chia. 2006. *RDF/XTM ontology construction based on a topic map-driven framework*. In Sugimoto, Shigeo, Hunter, Jane, Rauber, Andreas and Morishima, Atsuyuki, eds., *Digital Libraries: Achievements, Challenges and Opportunities 9th International Conference on Asian Digital Libraries, ICADL 2006, Kyoto, Japan, November 27-30, 2006. Proceedings*. Lecture Notes in Computer Science 4312: 523-6.

Yang, Huanhai and Sun Mingyu. 2013. Study on application of domain ontology in semantic information retrieval. *Applied mechanics and materials* 433-5: 1662-5.

Zhang, Jidong and Yu, Yisheng. 2006. Construction and research of ontology based on thesaurus. *Document, information and knowledge* 4: 82-5.

Zheng, Yu, Dou, Wenxiang, Wu, Gengfeng and Li, Xin. 2007. Automated Chinese domain ontology construction from text document. In Li, Kang, Fei, Minrui, William Irwin, George and Ma, Shiwei. *Bio-Inspired Computational Intelligence and Applications International Conference on Life System Modeling and Simulation, LSMS 2007, Shanghai, China, September 14-17, 2007*. Proceedings. Lecture notes in computer science 4688: 639-48.

Zeng, Marcia Lei, Panzer, Michael and Salaba, Athena. 2010. Expressing classification schemes with OWL 2 web ontology language. In Gnoli, Claudio and Mazzocchi, Fulvio, eds., Paradigms and Conceptual Systems in Knowledge Organization: Proceedings of the 2010 conference of the International Society for Knowledge Organization, Rome, February 2010. Advances in knowledge organization 12. Würzburg: Erlon Verlag, pp. 356-62.