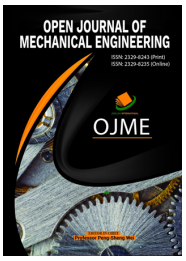
ZIBELINE INTERNATIONAL PUBLISHING

I : 676 687 (rint)
I : 676 67 (Online)

OD : O 8

Open Journal of Mechanical Engineering (OJME)

DOI : http://doi.org/10.26480/ojme.02.2016.0 .12



CrossMark

RESEARCH ON CONSTRUCTION METHOD OF DOMAIN ONTOLOGY BASED ON THESAURUS AND THEMATIC WORDS FOR HIGH-SPEED RAILWAY

Ziyu Liu[1,2*], Guang Yang[1,2], Bin Liu[1]

[1]Hebei University of Science and Technology
[2]State Key Laboratory for Manufacturing Systems Engineering Xi'an Jiaotong University
**Corresponding Author E-mail: purpleyuliu@sohu.com*

ARTICLE DETAILS

Article History:

Received 12 November 2017
Accepted 12 December 2017
Available online 1 January 2018

ABSTRACT

Domain ontology can be used to extend the user's retrieval words and index domanin knowledge. According to the actual situation of high-speed railway domain that is comprised by different professional fields, this paper puts forward a construction methodology of domain ontology based on thesaurus and thematic words of high-speed railway that is built by the experts. This paper details the construction principles for professional field ontology and the merger method for professional field ontologies. Guided by this method, we build various professional field ontology of high-speed railway, and then they are preliminary merged into one unified domain ontology of high-speed railway. Now the domain ontology of high-speed railway is used in semantic retrieval system for high-speed railway knowledge and it improve the recall and precision of system.

KEYWORDS

Arbuscular mycorrhizal fungus, root damage, mycorrhizoremediation

1. INTRODUCTION

In// Based on a study, ontology is playing an increasingly important role in the fields of software engineering, artificial intelligence, information retrieval and web service research [1]. According to the level of dependence on research fields ontology can be divided into Top Ontology, Domain Ontology, Task Ontology and Application Ontology [2]. Domain ontology can effectively organize the knowledge of that domain and make it easier to share and reuse.

High-speed railway (the new railway with speed above 250km/h) is the development mainstream of China. The High-speed railway development strategy of China is "introduction, digestion, absorption and innovate", so we should draw lesson from foreing advanced experience. In order to give facilities to query information about design, palnning, construction and operation of high- speed railway, the building of High-speed Railway Fundamental Information Database System is important. High-speed Railway Fundamental Information Database System (Z2006-094) is an important science and technology project of railway ministry of China. Under the support of this project and through a variety of ways to gather information, the number in document base of literature that is related with high-speed railway is more than thirty thousands. Because "thesaurus of railway" is not suitable for high-speed railway, so under the support of the project we organize experts to rebuild "thesaurus and thematic words of high-speed railway" which is apply to high- speed railway literature information. Among these the first class category is ten, the second category is 64, the third category is 208. 2488 subject headings Corresponds these categories. At the same time we also build "foundation data table of high-speed railway". In order to search needed information, we need a good search pattern. Now using "thesaurus and thematic words of high-speed railway", we indexed the literature in document base and relized keyword-based search model. However, such a model misses the actual semantic information of the text and it leads to the low recall and precision. In order to deal with this issue, ontologies are proposed for knowledge representation, which are nowadays the backbone of semantic web applications [3]. Semantic query expansion based on ontology can improve efficiently the recall and precision. So we will use it in our new search system for high-speed railway knowledge.

To achieve semantic retrieval, it needs to build domain ontology firstly. Most domains, for example, agriculture, aviation, railway, high-speed railway etc., have their own vocabulary, such as dictionaries, thesauri, Thesaurus, etc., so building the initial core ontology of industries domain based on the traditional classification/Thesaurus is a more scientific approach. Domain ontology can be used to extend the user's retrieval words and index domanin knowledge.

According to the actual situation of high-speed railway domain that is comprised by different professional fields, this paper puts forward a construction methodology of domain ontology based on thesaurus and thematic words of high-speed railway that is built by the experts. The rest of this paper is structured as follows. Section 2 describes the related works about ontology building. Section 3 considers methodology of ontology building for high-speed railway domain. In Section 4, guided by this method we build ontologies of high-speed railway. Finally, the main conclusions of our research, as well as the future work, are highlighted.

2. RELATED WORKS

There are two major methods for the construction of domain ontology: one is from the prespective of knowledge engineering to research the construction method of ontology that is called ontology engineering, another is transform the existing vocabulary resources directly to ontology that is called thesaurus-based ontology construction method.

2.1 Ontology engineering

Emphasize to build ontology according to certain norms and standard is the main feature of ontology engineering. Up to date, the following is the well-known typical ontology and methodology that occur in the development process of the ontology among ontology engineering: the enterprise ontology and Uschold & King method, The TOVE ontology and

Gruninger & Fox method, KACTUS and Bernaras method, CHEMICALS ontology and METHONTOLOGY method and The SENSUS ontology and method [4-7].

However, at present there is no standard and authority methodology in Ontology Engineering. Existing methodologies are born in specific projects and they are server for their projects. Due to different considerations of their respective areas and specific project, the process of build ontology is different. Based on a study, we can evaluate the maturity of methodology through the software life cycle method IEEE1074-1995 [8].

2.2 Thesaurus-based ontology construction

Many scholars presented to build ontology based on the existing thesaurus such as dictionaries, thesaurus and classification, and they have tried in practice. Thesaur have clear semantic structures, and can be used to extract concepts and relations. There are more than ten thesauri which have already been used to be converted into ontologies.

It can be divided into two categories according to the difference of construction method: one is to describe ontology directly based on thesaurus using the XML/RDFS syntax and does not adjust thesaurus. For example, Eman put forward to describe thesaurus using OWL that is one of the description language of ontology [9]. A researcher from China put forward to define descriptors using RDFS to realize the conversion from thesaurus to ontology; The other one is ontology thought, and it generate a new ontology through improving the thesaurus such as add or delete the concept of thesaurus, adjust the relations between concepts etc. [10]. For example, a researcher convert Agrovoc into agricultural ontology using automatic ontology learning system through extracting relations between concepts [11].

A group of researchers convert the controlled vocabulary in GEM into ontology use Ontoligua system [12]. Another researcher uses the controlled vocabularies of art and architecture thesaurus (AAT) to describe the ancient fruniture ontology [13].

Thesaurus and ontology are all build on the basis of comprehension of knowledge, so they have the premise of integration. But thesaurus and ontology are built for different objective. As a standardized vocabulary terms, thesaurus can improve the retrieval efficiency of computer. Building the relationship between concepts is the core of ontology, and the semantic content that the computer can understand is the premise of ontology. In order to convert thesaurus into ontology accurately, we need consider its characteristics, clean data and adjust semantic relation, so it will has some practical significance.

3. METHODOLOGY OF ONTOLOGY BUILDING FOR HIGH-SPEED RAILWAY DOMAIN

3.1 Methodology of Ontology Building for High-Speed Railway Domain

Some kinds of domains such as agriculture, railway, high-speed railway and aviation include different professional fields. For example, the high-speed railway domain consists of maintenance engineering, traction power supply, EMU and operation management etc. professional fields. The existing method of ontology construction is not suitable for high-speed railway domain that is composed by different professional fields.

The ontology of high-speed railway domain can be integrated by ontologies of the professional fields that are built on thesaurus and thematic words. Because “thesaurus of railway” is not suitable for high-speed railway, so under the support of railway ministry of China we organize experts to rebuild “thesaurus and thematic words of high-speed railway” which is apply to high- speed railway literature information. Among these the first class category is ten that includes maintenance engineering, traction power supply, EMU, signal and control, communication, operation management, safety and rescue, environmental engineering, transportation economy and comprehensive evaluation, the second category under the first class category is 64, the third category under the second category is 208. 2488 subject headings Corresponds these categories. At the same time we also build “foundation data table of high-speed railway” that includes 35 tables such as foundation data table for EMU, foundation data table for line and foundation data table for traction power supply etc.. The first class category of “thesaurus and thematic words of high-speed railway” includes the professional fields of high-speed railway, so we can build ontologies of the professional fields based on “thesaurus and thematic words of high-speed railway”. We can build property and instance of the professional fields based on “foundation data table of high-speed railway”.

We should clear that firstly the purpose of constructing high-speed railway domain ontology is to achieve semantic organization and semantic retrieval, secondly the hierarchical relation between concepts of domain ontology is selected based on “thesaurus and thematic words of high-speed railway” and use engineering thinking to construct domain ontology of high-speed railway. In the process of constructing ontology we should emphasize the participation of experts of high-speed railway domain. Because even if construct ontology through engineering method, it also need expert to identify and evaluate.

According to the actual situation of high-speed railway domain and based on the two major methods ontology engineering and thesaurus-based ontology construction, this paper puts forward a construction methodology of domain ontology that is composed of multi-disciplinary based on thesaurus and thematic words of high-speed railway that is built by the experts, shown as Fig. 1. First we build various majors ontology of high-speed railway, and then they are preliminary merged into one unified domain ontology of high-speed railway. Building ontology of the professional field and merging them into one are important, the next following details all parts of this two steps.

3.2 Principle of constructing ontology for the professional field

The model of ontology main includes concepts, properties, relations and instances [14].

(1) Choice concept

The first class concept main comes from the words of “thesaurus and thematic words of high- speed railway”. The second class concept is the classification of the first class concept, and if the first class concept has two or more division method, we often choice the most general division method as the second class concept. If the second class concept can still classified, divided down, until no classification.

For example, EMU is the first class concept of the ontology of EMU professional field. EMU has three division methods: according to the dynamic configuration, according to the fashion of supply power and according to speed grade. Among these the most commonly used is according to the fashion of supply power in which EMU is divided into “EMU of power distributed” and “EMU of central power”. So the second class concept of EMU is “EMU of power distributed” and “EMU of central power”.

(2) Determine the property

The property of the professional field ontology main comes from “foundation data table of high- speed railway”, and if the properties of the concept is not complete, that need the experts of high- speed railway domain to add.

For example, we can add “type”, “manufacture”, “country” and “starting acceleration” etc. properties to EMU concept according to “foundation data table of high-speed railway”.

(3) Determine the instance

The instance in “foundation data table of high-speed railway” is not complete, so it needs domain experts to add. If the concept has sub-concept, the instance is added to the bottom concept.

For example, we can add “CRH1”, “CRH2”, “CRH3” etc. instances to “EMU of power distributed” according to “foundation data table of high-speed railway”.

(4) Determine the relation

There are two types of relations in ontology. One is hierarchical relationship including the hyponymy relation and instance relation and it can get from “thesaurus and thematic words of high-speed railway”. The other is non hierarchical relationship, and it can be described by adding the special property of concept to connect two concepts. Non hierarchical relationship need to be determined by domain experts. For example, add the special property of “supply power” to the concept of “pantograph”,

which can describe the non hierarchical relationship between the concepts of "pantograph" and "EMU". Therefore, we can get the relation: pantograph <supply power> EMU.

3.3 Merge the professional field ontologies

When the professional fields's core ontologies are built, we can merge the professional field ontology according to the relationship between different professional fields into one unified domain ontology of high-speed railway. For example, the concept of "contact network system" in traction power supply ontology associate with the concept of "pantograph" in EMU ontology through the relation of <supply power>.

4. BUILDING ONTOLOGIES FOR HIGH-SPEED RAILWAY

Building ontology demands standardized expression fashion and work steps. Describing language and modeling mode are chosen according to the level of ontology, the use of ontology, the principle of building ontology and evaluation criteria. In our system we use OWL describing language for ontology. OWL is a standard of ontology describing language in semantic web that is recommended by W3C, and it is an extension on the basis of RDFS [15]. We adopt Protégé that is developed by Stanford University to build our domain ontologies. Method about building ontology is not mature, and there is not has a complete and unified methodology. However, a number of typical ontology and methodology have been developed. In our system we adopt our method proposed in section 3 to build various professional fields ontology, and they are preliminary merged into one unified domain ontology of high-speed railway. At present, we have constructed five professional fields's core ontology of EMU, operation management, safety and rescue, traction power supply and maintenance engineering. The thesaurus and thematic words of EMU includes four hundred and twenty one words. The depth of the EMU ontology network is 3. By now, we have defined forty five concepts, sixty and six properties, six user-defined relations and one hundred and ten instances. The thesaurus and thematic words of operation management includes one hundred and five words. The depth of the operation management ontology network is 3. By now, we have defined forty and eight concepts, thirty properties, one user-defined relation and twenty nine instances. The thesaurus and thematic words of safety and rescue includes thirty two words. The depth of the safety and rescue ontology network is 2. By now, we have defined nineteen concepts, seven properties, one user-defined relation and six instances. The thesaurus and thematic words of traction power supply includes three hundred and seventy eight words. The depth of the safety and rescue ontology network is 4. By now, we have defined three hundred and fivety six concepts, ten properties, five user-defined relations and thirty five instances. The thesaurus and thematic words of maintenance engineering includes nine hundred and eighty nine words. The depth of the safety and rescue ontology network is 4. By now, we have defined nine hundred and twenty concepts, twenty five properties, one user- defined relation and fourteen instances.

Figure 1: The flow of domain ontology construction based on thesaurus and thematic words of high-speed railway

Figure 2: The segment of high-speed railway domain ontology

Use Protégé , they are preliminary merged into one unified domain ontology of high-speed railway. The depth of high-speed railway domain ontology network is 4. Fig. 2 is a segment of high-speed railway domain ontology.

5. CONCLUSION

Semantic web on ontology-based provides a new guidance way for information retrieval. Domain ontology can be used to extend the user's retrieval words and index domain knowledge. According to the actual situation of high-speed railway domain that is comprised by different professional fields, this paper puts forward a construction methodology of domain ontology based on thesaurus and thematic words of high-speed railway that is built by the experts. Guided by this method, we build various majors ontology of high-speed railway, and then they are preliminary merged into one unified domain ontology of high-speed railway. Under which we can achieve semantci retrieval for domain knowledge. Now the domain ontology of high-speed railway is used in semantic retrieval system for high-speed railway knowledge and it improve the recall and precision of system. In our future work, we will perfect the domain ontology and consider the evolution problem of domain ontology of high-speed railway.

ACKNOWLEDGMENTS

This work was financially supported by the Natural Science Foundation of Hebei Province of China (F2012208018, G2010000903), the project of Natural Science Foundation of China (71271076), the social science fondation of Hebei Province of China (HB12GL040), the project of State Key Laboratory for Manufacturing Systems Engineering (Xi'an Jiaotong University) (sklms2011002), the science and technology research and developing plan project of railway ministry of China (Z2006-094) under which the present work was possible.

REFERENCES

[1] Klein, M., Bernstein, A. 2001. Searching services on the semantic: Web using process ontologies, Proceedings of the International Semantic Web Working Symposium (SWWS), Amster2, dam, IOS Press, 159-172.

[2] Zhi, J. 2001. Ontology research in knowledge engineering, Knowledge engineering and knowledge science, Bejing, Tsinghua University press, 447-465.

[3] Soner, K., Ozgur, A., Orkunt, S., Samet, A. 2011. An ontology-based retrieval system using semantic indexing, Information Systems, In press, Corrected Proof, 1-12.

[4] Uschold, M., Gruninger, M. 1996. Ontologies: principles, methods, and applications, Knowledge Engineering Review, 11, (2), 93-155.

[5] Bernaras, A., Laresgoiti, I., Corera, J. 1996. Building and reusing ontologies for electrical network applications, in: Proc of the European Conf on Artificial Intelligence, Budapest, Hungary, John Wiley and Sons,

298-302.

[6] Fernandez-Lopez, M., Gomez-Perez, A., Pazos-Sierra, A. 1999. Building a chemical ontology using methontology and the ontology design environment, IEEE Intelligent Systems and Their Applications, 14, (1), 37-46.

[7] Knight, K. 1995. Filling knowledge gaps in a broad-coverage machine translation system, In: The 14th Int'1 Joint Conf on Artificial Intelligence, Montreal, Canada, 1390-1396.

[8] Li, J., Meng, L.S. 2004. Comparison of seven approaches in constructing ontology, New Technology of Library and Information Service, 112, (7), 17-22.

[9] Eman, J. 2005. Owl Exports from a Full Thesaurus, Bulletin of the American Society for Information Science and Technology, 32, (1), 22-26.

[10] Mao, J. 2003. Research on RDF-based thesaurus, Journal of the China Society for Scientific and Technical Information, 22, (2), 163-168.

[11] Asanee, K. 2006. Automatic Term Relationship Cleaning and Refinement for AGROVOC [OL].

[12] Qin, J., Paling, S. 2001. Converting a controlled vocabulary into an ontology: the case of GEM, Information research, 6, (2), http://informationr.net/ir/6- 2/paper94.html.

[13] Wiefinga, B.J., Schreiber, A., Wielemaker, J., Sandberg, J.A.C. 2008. From Thesaurus to Ontology, In: Proceedings of the 1st international conference on knowledge capture, 194-201, http://www.cs.vu.nl/guus/papers/Wielinga01a.pdf.

[14] Tang, A,M., Zhen, Z., Fan, J. 2005. Thesaurus-based approach to build domain ontology, New Technology of Library and Information Service, 122, (4), 1-5.

[15] Kong, T.Y., Li, W.L., Zhang, H.O. 2008. Research in medicine ontology-based information retrieva1 system, Journal of Hebei University of Science and Technology, 29, (3), 223-226.