A Domain Ontology Construction Method with Ontology Modification Effort Assessment

Yuehua Yang, Yuan Ping, Junping Du and Hui Ma

1 Introduction

With the rapid development of semantic Web technology, ontology technology has been widely concerned [1, 2]. For specific domains, most existing ontology lack domain words and semantic relationships between them. When these ontology are applied to specific domains, the application effect is not ideal. Therefore, in many cases, it is necessary to construct new ontology according to the situation of the domain itself [3, 4]. Therefore, it is essential to study and propose an effective domain ontology construction method.

In the construction of domain ontology, successful projects mostly refer to the software engineering approach, and some domain ontology construction methods are summed up according to the construction process. But these methods don't consider fully, and seldom consider the ontology modification efforts assessment [5, 6]. So they are not always applicable to other domain or applications.

In view of the above problems, in this paper a domain ontology construction method (DOCM) is proposed. This method divides the ontology construction process into ontology requirements analysis, domain knowledge analysis, ontology

Y. Yang (✉) · Y. Ping (✉) · H. Ma
School of Information Engineering, Xuchang University, Xuchang 461000, China
e-mail: yyh0504@126.com

Y. Ping
e-mail: pyuan.lhn@xcu.edu.cn

H. Ma
e-mail: xcmahui@qq.com

J. Du
Beijing Key Laboratory of Intelligent Telecommunications Software and Multimedia, School of Computer Science, Beijing University of Posts and Telecommunications, Beijing 100876, China
e-mail: junpingdu@126.com

Z. Deng (ed.), *Proceedings of 2017 Chinese Intelligent Automation Conference*, Lecture Notes in Electrical Engineering 458,
https://doi.org/10.1007/978-981-10-6445-6_52

establishment, ontology evaluation, and ontology modification effort assessment. In order to determine whether it is necessary to modify the ontology, an evaluation method based on ontology modification effort and elements importance is proposed. When updating the ontology, we can use this method to evaluate the modification effort to avoid the loss outweighs the gain in ontology use and modification.

2 DOCM

In recent years, a large number of domain ontology has been established, but few is really used in the practical applications. The problems mainly include: requirement analysis is not sufficient; ontology construction process is completed very casually; ontology is rarely evaluated and modified; the usability of the constructed ontology is generally poor. In order to ensure that the ontology construction process is relatively normal, and the ontology is easy to expand, a new domain ontology construction method DOCM is proposed. The framework of DOCM is shown in Fig. 1.

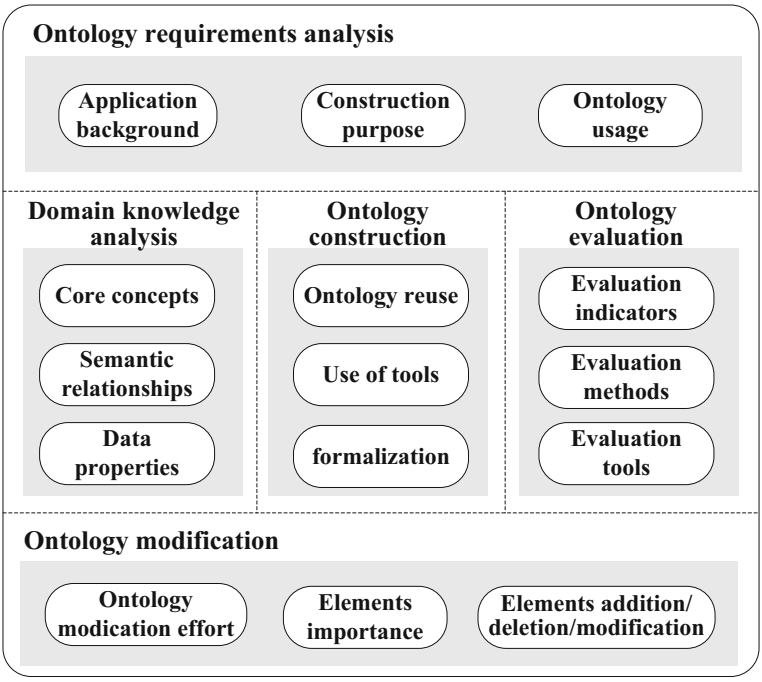

Fig. 1 Framework of DOCM

(1) Ontology requirements analysis: to understand and define the specific application background, construction purpose and usage of ontology. In order to combine the construction requirements of ontology and specific applications tightly, we need to have a certain understanding of the specific functional requirements of the application itself. Many following steps of ontology construction are corresponding to the requirements. Therefore, the more detailed the requirements description of ontology is, the clearer the content that the domain ontology should contain and the following construction steps are.
For example, an emergency domain ontology is constructed to provide semantic support for emergency information retrieval. Users' queries can be optimized with the rich semantic relationships and reasoning mechanism of the emergency domain ontology, thus making up for the shortcoming that users' queries expressions are not sufficient. The ontology can also be used to index the documents, and filter out the language ambiguity of the text when extracting the features of the text content.

(2) Domain knowledge analysis: to analyze and extract the knowledge based on the application requirements, such as core concepts, semantic relationships and data properties. After all, the understanding of domain knowledge from different perspectives is different, and the application demands are diversified.
In emergency information retrieval, the domain ontology needs to include more comprehensive emergency knowledge so that it can provide better semantic support. After analyzing the emergency domain knowledge, we can conclude that it need to include the core concepts about the classes of emergencies, the emergency evolution process, the emergency actions and decisions; semantic relationships between emergency concepts, such as classification, causality, coupling and sequence relationships. etc.; data properties that describe the concepts, for example, emergencies can be described with properties such as time, place, level, hazard, controllability, victim, influence scope, nature, characteristics, and precautionary methods, so these data properties need to be added into the ontology.

(3) Ontology construction: to use ontology building tools to add domain concepts, data properties, semantic relationships between the concepts and axioms for the domain ontology, or use a formal language to describe the content clearly. Besides reuse and integration of the existing ontology are also need to do. For example, the emergency domain ontology can be built by using Protégé tool and formalized in OWL language. Then it can be applied into emergency information retrieval.

(4) Ontology evaluation: to evaluate the quality of ontology after it is constructed to ensure that ontology meets the demands. It can also help the user to quickly find the appropriate ontology to support the application, so as to ensure the quality of the application system. In the evaluation process we need to select the appropriate evaluation indicators and evaluation methods or evaluation tools.
For example, OntoQA method proposed by LSDIS laboratory of computer science department of Georgian University is used to evaluate emergencies

domain ontology. The classes richness, the properties richness, the relationship richness, the inheritance richness and the average distribution of instances are selected as the evaluation indicators to evaluate the ontology quality. According to the indicators calculation results we can know the quality of the emergency domain ontology to a great extent.

(5) Ontology modification: to modify the ontology appropriately in the practical application, such as add new concepts, modify or delete the inaccurate concepts in the ontology. Before modification we should evaluate the ontology modification effort and the importance of the elements to be modified. The importance of the assessment, because there are usually some associations between ontology elements, changing an element will affect other elements. If the modification effort is too large, modification is not so necessary.

3 Ontology Modification Effort Assessment Method

The number of concepts, properties and semantic relationships between concepts, which are the three main elements of ontology, will affect the complexity of an ontology comprehensively. Only considering one of the elements is not sufficient to fully assess the efforts of modifying the ontology. The more concepts covered in the ontology, the richer the entity knowledge ontology can express is. Properties can describe some internal features of the concepts. The number of properties indicates the completeness and complexity of the concepts in an ontology. The more the relationships, the more work to do when modify a concept, because the associated sub classes or parent classes need to be checked.

Therefore, a method used to evaluate the effort of modifying a given ontology is presented. This method considers the concepts, properties and semantic relationships in modification effort assessment. Here, the number of concepts is represented by c, the number of semantic relationships is represented by r, and the number of properties is represented by p, hence forming a point (c, r, p). So this method is also called a three-dimensional modification assessment method.

For a given ontology with the higher complexity, the farther the point (c, r, p) is away from the origin (0, 0, 0), the more the modification effort is, and vice versa. Similarly, after adding new elements to the ontology, the farther the corresponding point is away from point (c, r, p), the more the modification effort is. In addition, the more complex the given ontology, that is, the ontology contains more concepts, properties, or relationships, the more the effort of adding new elements is. So in order to better measure the ontology modification effort, the case above should also be considered. Thus the following formula is proposed to evaluate the effort of adding new elements to the ontology:

$$e = k\sqrt{(c_2 - c_1)^2 + (r_2 - r_1)^2 + (p_2 - p_1)^2} \quad (k = c_1 + r_1 + p_1 + 1) \qquad (1)$$

In formula (1), c_2, r_2, p_2 represent the number of concepts, relationships and properties respectively after adding new ontology elements. c_1, r_1 and p_1 represent the number of concepts, relationships and properties of the original ontology. k is a coefficient. It is used to ensure that when adding the same number of elements into the ontology, the more the original elements, the greater the effort estimate of adding new elements.

For example, a domain ontology originally contains 50 concepts, 25 semantic relationships, and 8 properties. Through evolution, more domain concepts, relationships and properties can be obtained. If the number of concepts, relationships and properties is increased to 55, 28, 10 respectively, according to the formula (1) the ontology modification effort estimate is:

$$e_1 = (50 + 25 + 8 + 1) \times \sqrt{(55 - 50)^2 + (28 - 25)^2 + (10 - 8)^2} = 517.8$$

If the concepts, relationships and properties are increased to 60, 31, 12 respectively, according to the formula (1) the modification effort estimate of the ontology is:

$$e_2 = (60 + 31 + 12 + 1) \times \sqrt{(60 - 50)^2 + (31 - 25)^2 + (12 - 8)^2} = 1282.2$$

According to e_1 and e_2 we can understand the ontology modification effort and complexity that is measured based on the three factors: the number of concepts, the number of relationships and the number of properties.

If the concepts, relationships and properties are increased from 50, 25, 8 to 55, 28, 10 respectively, the concepts, relationships and properties are increased by 5, 3 and 2 respectively, the ontological modification effort estimate e_1 is 511.6. On this basis the concepts, relationships and properties are increased to 60, 31, 12 respectively, also increased by 5, 3, 2. According to formula (1) the modification effort estimate e_3 is:

$$e_3 = (55 + 28 + 10 + 1) \times \sqrt{(60 - 55)^2 + (31 - 28)^2 + (12 - 10)^2} = 579.5$$

As the results show, $e_3 > e_1$. This indicates that the more the original ontology concepts, relationships and properties, the more the modification effort. Thus, a three-dimensional modification assessment method can also be used to assess whether a given ontology is worth modifying for reuse. In addition, the three-dimensional modification assessment method plays an important role in many other aspects of the ontology. In order to clarify some actual situations, the following tests and comparison are done on some real ontology. These ontology are very large, so we will need to do a lot of work to modify them. The results is shown in Table 1.

Table 1 Results of testing on real world ontologies

Ontology name	No. of concepts	No. of relationships	No. of properties	Modification effort estimate
BCGO	1882	117	3644	2.3×10^7
PO	1691	10	2687	1.4×10^7
SO	936	1	840	2.2×10^6
EO	906	90	1114	3.0×10^6
HDO	8946	15	6918	1.8×10^8

Results from Table 1 on these real world ontology show that using the three dimensions ontology modification effort assessment method can assist in providing general understanding about an ontology. That is, when using the proposed method to calculate the modification effort estimate results, ontology modification effort and complexity can be estimated. This can be significantly beneficial when a decision is about to be made concerning an ontology. If modifying the ontology need more effort, we should consider to divide it into several ontology.

Assume that in some scenarios ontology about "Places" is needed. And the ontology SWETO covers the concepts, relationships and properties of some topics including person, places, academic department, event, organization, etc. So SWETO ontology offers the modeling that is needed about "Places". Table 2 shows the modification effort estimate before and after extracting the concepts and relationships about place from SWETO. Before applying the proposed method, it is a little vague that how detailed this ontology is, and what these numbers mean.

It can be seen from Table 2 that the modification effort estimate before extracting the concepts and relationships about place is large. Therefore, after considering the modification effort estimate, the expected effort can be grasped when using SWETO ontology, so that the place module can be independently selected and extracted. To modularize the ontology, we can use a Web-based OWL extractor to separate an ontology based on user-specified feature words and options. After extracting the place module we calculate the modification effort estimate again with formula (1) and find that the estimate becomes smaller. And the relevant concepts and relationships of place become relatively rough.

The three-dimensional modification effort assessment method can also be used when the ontology owner needs to adjust the ontology size in order to ensure that the ontology is easy to understand and update. Through this assessment method, the ontology owner can track the ontology development process. In order to ensure that the ontology modification effort estimate is small, we can perform frequent tests based on this assessment method when updating the ontology, and then determine whether new elements are all added to the ontology. If not, we need to determine which new elements should be retained, which is mainly determined by the

Table 2 Modification effort estimate before and after testing

Ontology	No. of concepts	No. of relationships	No. of properties	Modification effort estimate
Before	114	69	111	5.1×10^4
After	8	14	7	5.3×10^2

Table 3 The importance

Candidate ontology elements	Weighted percentage	Importance (%)
Fire hazard	3.91	49.06
Fire control	1.13	14.18
Scene	0.95	11.92
Cause	0.82	10.29
Casualty	0.43	5.40
Fire behavior	0.40	5.02
Accident	0.33	4.14
Total	7.97	100

importance of the elements. For a set of elements (assuming the number of the elements is n), the importance of each element c is determined according to the following formula:

$$I_c = \frac{w_c}{\sum_{i=1}^{n} w_i} \tag{2}$$

where I_c is the importance of concept c in the group; w_c is the weight of concept c among all the concepts while the denominator represents the sum of weights of the concepts in the group. The weighted percentage of each word has been identified by the content analysis software NVivo (http://www.qsrinternational.com/) comparing the frequency of each word in the corpus of data with other words.

For example, use "query"—"word frequency" function of NVivo software to obtain the weighted percentage of a group of ontology elements from the corpus, as shown in Table 3, and then calculate the importance of each element according to the formula (2).

4 Conclusion

A new domain ontology construction method DOCM is proposed to perfect the domain ontology construction process provide semantic support for multiple domain applications. In addition to the ontology requirements analysis, domain knowledge analysis, ontology establishment and ontology evaluation, ontology

modification effort assessment is also included in this method to determine whether it is worth modifying the ontology. And a three-dimensional modification assessment method is proposed. In the modification effort assessment the changes of three main elements of ontology including concepts, properties and the semantic relationships between concepts are all considered. If the modification effort estimate is too large, the ontology will be no longer modified. Therefore, based on the assessment method, the ontology owner can track the ontology development process. Besides, the knowledge in the ontology can be extracted and reused so that the effort estimate for ontology modification and use can be reduced. Domain ontology constructed with the proposed method will include more domain terms and semantic relationships and can also get better application effect when applied to the domain applications.

Acknowledgements This work was supported by National Natural Science Foundation of China (No. 61320106006), Foundation of He'nan Educational Committee (No. 17A520013, No. 15A520096, 18A520047), Program for Science &Technology Innovation Talents in Universities of He'nan Province (18HASTIT022), Foundation for University Key Teacher of He'nan Province (No. 2016GGJS-141), Outstanding Young Teacher Project of Xuchang University, and Xuchang Science and Technology Development Program (No. 1504017).

References

1. Yang YH, Du JP, Ping Y (2015) Ontology-based intelligent information retrieval system. J Softw 26(7):1675–1687
2. Zhitomirsky-Geffet M, Erez ES, Judit BI (2017) Toward multiviewpoint ontology construction by collaboration of non-experts and crowd sourcing: the case of the effect of diet on health. J Assoc Inf Sci Technol 68(3):681–694
3. Liu LZ, Zhao XL, Wang HS et al (2015) Constructing domain affective ontology based on product features. Trans Beijing Inst Technol 35(5):538–544
4. Azam E, FarookhKhadeer H, Elizabeth C (2013) University social responsibility ontology. Eng Intell Syst 21(4):271–281
5. Yu XL, Wang XP (2016) A construction method of emergency domain ontology based on scenario. ICIC Express Lett B Appl 7(4):947–953
6. Abdulelah AA, Austin CM (2016) Three dimensions ontology modification matrix. In: International conference on information management, pp 77–83