

Studies on ontology-based TCM knowledge base construction mode

Chunhua Liao¹, Yanqiu He², Guangyao Xiong^{3*}

¹School of Computer Science, Jiangxi University of Traditional Chinese Medicine, Nanchang, 330004, China

²College of Humanities, JiangXi University of Traditional Chinese Medicine, NanChang, 330029, China

³Network Education Technology Center, Jiangxi University of Traditional Chinese Medicine, NanChang, 33000, China

*Corresponding author

Keywords: Ontology, Knowledge base, Ontology construction, TCM.

Abstract. This article has researched the model based on ontology construction domain knowledge base system, and provided uniform description and expression of knowledge base model by using BNF (Backus-Naur Form). Based on this, the article has studied and constructed a knowledge base system based on TCM domain ontology, and proposed methods of constructing ontology-based TCM knowledge base.

Introduction

Human beings have truly elevated knowledge from concept to knowledge science by using computer technology during the process of stepping information-based society to knowledge-based society [1]. However, with the advancement of technology, the problem that there is not a uniform standard for knowledge representation is getting increasingly serious, which has influenced knowledge sharing and reusing [2][3].

In 1977, B. A. Feigenbaum, a computer scientist in Stanford University proposed the concept of KE (Knowledge Engineering) at the 5th International Joint Conference on Artificial Intelligence. He believed that "KE is the principle and method of AI application, as well as the means of providing solutions for application puzzles required for expertise knowledge. Appropriate using the composition and explanation of expertise knowledge acquisition, representation and reasoning is an important technical issue for knowledge-based system design" [4]. KE proposed by Guus Schreiber is a modelling activity with model as a purposeful abstraction on a certain part of reality and modelling as a good description established from a few aspects of knowledge but other aspects not into consideration" [5]. Thus it can be seen that knowledge representation and knowledge application based on this are most principal research contents in KE domain. Knowledge model is a well-structured knowledge representation and knowledge can be described and processed by using knowledge model [1].

Knowledge representation and knowledge ontology

Knowledge model

The core task of KE is effective knowledge representation. Traditional knowledge representation methods include first-order predicate logic representation, production-type representation, Petri net representation and OO (object-oriented) representation, etc. They can be roughly categorized into symbol-based representation and connectionism-based representation [6]. There exist some disadvantages of traditional knowledge representation methods during the process of knowledge representation, expression and sharing. For instance, they cannot guarantee the uniqueness and unambiguousness of knowledge comprehension during knowledge transmission and sharing. Also in complicated knowledge representation and reasoning, excessive scale of system atomic knowledge will result in combination explosion. In such a case, conflict resolution ability, the size, and amount of

knowledge granularity and organization form would directly determine the availability of knowledge representation methods. The concept of knowledge ontology has been introduced to KE domain so as to solve abovementioned issues^[6]. Knowledge ontology is the specification of related knowledge sharing concept model and explicit formalization^[7], and it has provided basic terminologies (knowledge atom) and relations in knowledge model, and used these terminologies and relations to constitute the extension rules and complex definition of knowledge^[8]. Introducing ontology to knowledge base^[9] can formalize the representation of domain knowledge concepts and relationship between concepts for a structured and formalized knowledge representation. Ontology-represented knowledge can be independent of knowledge base application system, which makes knowledge reused between different knowledge base systems. It can help improve information sharing and transmission of knowledge base and help more convenient communication between researchers in different knowledge domains. Due to the support of separation between operation and representation from ontology, ontologies in different domains can apply same operation processing knowledge. Based on the modelling of ontology knowledge model, knowledge workers can conveniently establish domain ontology by consummating concepts, rules, relations and examples involved in model and operate in an easy way.

Formalized definition of knowledge ontology can be denoted by a tetrad [6]:

$$KO ::= \langle MD \rangle \langle KA \rangle \langle Rel \rangle \langle Rul \rangle \quad (1)$$

Hereinto, MD (Meta-Data) refers to the aggregate of several knowledge attributes related to knowledge ontology such as its name, creation time and version, etc. KA refers to knowledge atom, as the min unit of knowledge representation unit in knowledge model such as axiom, concept and basic operational relationship. Its formalized definition is:

$$KA ::= \{k_i | 1 \leq i \leq n, k_i \notin \phi, k_i \in \Omega\} \quad (2)$$

Hereinto, Ω refers to domain knowledge field, k_i is the knowledge atom of knowledge universe; Rel refers to the aggregate of correlation, effect and influence between knowledge atoms and between different knowledge entities constituted by knowledge atoms. Its formalized definition can be denoted as:

$$Rel ::= \{rij(k_i, k_j) \vee rim(k_i, M(k_1, k_2, \dots, k_m))\} \quad (3)$$

Also satisfy $\{1 \leq i, j, m \leq n, k_i, k_j, k_m \in KA, ri \notin \phi\}$, hereinto, $rij(k_i, k_j)$ refers to the relationship between knowledge atom k_i and knowledge atom k_j , $rim(k_i, M(k_1, k_2, \dots, k_m))$ refer to the relationship between knowledge atom k_i and knowledge entity $M(k_1, k_2, \dots, k_m)$. $M(k_1, k_2, \dots, k_m)$ refers to a knowledge entity with larger granularity constituted by knowledge atoms through rule combination. Its formalized definition can be denoted as:

$$M(k_1, k_2, \dots, k_m) ::= \{\Sigma k_i k_j \vee \Pi k_i k_j | 1 \leq i, j \leq n; k_i, k_j \in \Omega\} \quad (4)$$

Rul is the aggregate of some operations and rules after the combination of relationships between knowledge atoms or knowledge entities.

Knowledge ontology

Perez, etc. [10] concluded knowledge entity in a certain knowledge domain into five fundamental elements, Class, Relation, Function, Axiom and Instances. We hereby apply BNF (Backus-Naur Form) to provide description of knowledge ontology as below:

```

<knowledge Ontology> ::= <Meta-data> <Class> | <Object>
<Meta-data> ::= <Md-id> <Md-name> <Class>
<Class> ::= <Instances> { <Instances> [ <Relation> ] }
<Instances> ::= <kd-entity> { <kd-atom> }
<Relation> ::= <sys-type> | <user-type>
< sys-type > ::= <part-of> | <instance-of> | <attribute-of>
<kd-entity> ::= <kd-atom> { < kd-atom > }
<kd-atom> ::= <axiom> | < conception> | <principle> | <Terminology>
    
```

Formalized description of domain knowledge can ensure the uniqueness and accuracy of knowledge comprehension in knowledge transmission process, and improve knowledge ontology sharing and reuse. The hierarchical relationship between knowledge atom and knowledge entity in

knowledge ontology can well reflect the layered and constraint relationship between knowledge such as concept, axiom and relation, control the size and amount of knowledge granularity and effectively solved the combination explosion issue in knowledge representation and reasoning, and strengthened the usability of representation method of knowledge ontology.

Ontology-based knowledge model

Knowledge model means to give a formalized description of knowledge content and knowledge operation. A knowledge model can provide normative instructions of data and knowledge structure as required. A knowledge model includes three parts, domain knowledge, inference knowledge, and task knowledge. Each part contain a group of relevant knowledge structure, i.e. knowledge category [5]: (1) domain knowledge is the detailed description of domain-specific knowledge and information type discussed in an application; (2) inference knowledge, also called method knowledge, is the description of basic inference procedures of using domain knowledge, i.e. fundamental components of inference machine or inference engine. Inference knowledge defines the model of these fundamental components; (3) task knowledge describes the goal of application, and how to achieve the goal through task breakdown into sub-tasks and inference.

Corresponding to knowledge representation system, domain knowledge model belongs to "fact/concept" layer, inference/method knowledge model belongs to "rule" layer. Task knowledge model has described application task. All knowledge models must be established based on knowledge ontology, i.e. to satisfy the definition mode of abovementioned knowledge ontology for a uniform description and representation of knowledge model by means of BNF after combining the components of knowledge model with the primitive of knowledge ontology [1]. Its knowledge model framework is shown as below:

- (2) Domain knowledge model representation
- (3) Inference/method knowledge model representation
- (4) Task knowledge model representation

Studies on TCM knowledge base structure

Principles of TCM ontology construction

There is yet no final verdict on principles of ontology construction. In 1995, T.R. Gruber proposed the five most popular principles for ontology construction guidance.

(1) Specificity: ontology must validly explain the definition term, i.e. the ontology must give clear, authoritative, accurate and standard description of concepts involved. Definition of concept must be independent, objective and authoritative, combined with specific domain or specialty background in a formalized description by logic axioms.

(2) Consistency: the ontology must be consistent as a whole, i.e. the ontology supports inferences in accord with its definition with correct concept definition inferred by axioms.

(3) Extendibility: the ontology must provide concept basis for predicable tasks, i.e. ontology design should be continuously adjusted and expanded with the change and development of domain. Ontology design should be supportive to define new concepts based on existing concepts to satisfy special demands.

(4) Minimal coding deviation: concept description should not depend on specific symbol layer representation method, i.e. concept description is free from the influence of symbolic coding.

(5) Minimal ontological commitment: ontological commitment should be the minimum, i.e. the ontology can satisfy the given knowledge sharing demand. The oversized coverage area of ontology may result in inconformity of concept definition and ambiguity.

Construction method and process of TCM knowledge ontology

During constructing TCM ontology knowledge base, we have used Stanford University's seven-step method for TCM knowledge ontology construction. 1) Confirm the domain of ontology; 2)

Consider to reuse existing ontology; 3) List important terminologies in ontology; 4) Feasible methods to define hierarchy system between classes and commutate hierarchy system include: top-down method, bottom-up method and comprehensive method; 5) Define the attribute of class; 6) Define the value range of attribute; 7) Create instances.

Besides, knowledge base cannot be constructed without the participation of domain experts. Thus we have invited TCM experts to inspect the whole process to complete the work in close cooperation with knowledge engineers. Process of constructing TCM knowledge ontology is as below:

1. Demand analysis: firstly, obtain relevant classes, subclasses and instances TCM domain for conceptualization. Then confirm the hyponymy between concepts. Specific task is to clarify the range, purpose and use of ontology construction.

2. Domain ontology reconstitution: confirm basic requirements for domain ontology reconstitution based on concepts. Effectively reused established system concepts and correlations for supplement and perfection. Ontology integration can improve the efficiency of domain ontology construction so as to save the period and cost of ontology construction.

4. Ontology analysis: define the concept of TCM and layer concept system. TCM knowledge classification should be based on TCM dialectical system. Establish TCM knowledge ontology and create subclasses based on TCM theoretical knowledge under the guidance of experts in TCM domain.

5. Ontology realization: define TCM knowledge attributes: composition attribute, medicine name, medicinal parts, preparation, dosage form, medicine properties, efficacy attribute, etc.

Ontology-based TCM knowledge base system

Ontology-based TCM knowledge base model is shown in diagram 1. There exist three types of personnel in TCM ontology knowledge base system, who are target user groups of the system, TCM worker, knowledge engineer and TCM expert. Different users (roles) have different limits of authority.

Procedures to construct ontology-based TCM ontology are as below: (1) construct TCM knowledge ontology model based on TCM subject contents; (2) edit TCM ontology based on knowledge ontology model; (3) conduct knowledge inspection based on concepts, axioms and interrelations in TCM domain; (4) generate TCM domain knowledge ontology and establish knowledge base; (5) develop man-machine interacted ontology editing interface. Specific construction process is as below: knowledge engineer identifies knowledge such as concepts of TCM domain, relationship between concepts and knowledge inference rules, and defines the relationship such as father-son relationship, member relationship, functional similarity or contrary relationship, leading-subsequent relationship, etc. Then extract concepts and other entity attributes. Common relationship such as transitive relation, symmetric relation and function relationship and inverse relationship can be obtained through analyzing the relationship between concepts in TCM domain, between concepts and instances, and between different instances in TCM domain.

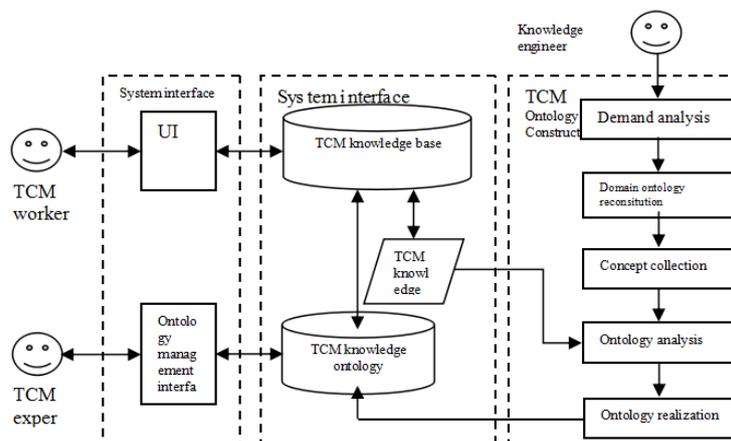


Fig. 1. Chinese medicine ontology knowledge base model

TCM experts can add new concepts and relations through ontology editor interface provided by the system. Ordinary TCM workers can obtain TCM ontology on ontology display interface to learn TCM knowledge.

Work in procedure (1), (2) and (3) will be completed by knowledge engineers under the support of TCM experts. Work in procedure (5) will be completed by computer processing or man-machine interaction.

Conclusion

Traditional knowledge base systems such as subject headings and thesaurus have limited representation ability with obvious limitation in complex knowledge representation. Also there may exist combination explosion problem of ambiguity in knowledge comprehension and knowledge operation during knowledge transmission. This article tries to introduce ontology to TCM knowledge base system so as to overcome abovementioned problems. Based on researching traditional knowledge representation methods, this article has used the formalized definition of knowledge ontology of $\langle MD \rangle \langle KA \rangle \langle Rel \rangle \langle Rul \rangle$ tetrad^[6], and give formalized description of domain ontology based knowledge base model by using BNF and combining fundamental modelling elements of domain ontology. At last, the article has researched implementation methods to construct TCM domain knowledge base based on this knowledge base model.

Research shows that introducing ontology and giving a formalized explanation of concepts and their relations can guarantee the uniqueness and accuracy of knowledge comprehension during the process of transmitting domain knowledge, and can improve knowledge ontology sharing and reuse. In addition, the hierarchical relationship between knowledge atom and other knowledge entities in knowledge ontology has well reflected the hierarchical and constrained relationship between concepts, axioms and relations in the domain. Formalized definition specification can map production or restrain new knowledge rules, control the size and amount of knowledge granularity, and effectively solve the combination explosion problem in knowledge representation and reasoning, and strengthen the usability of knowledge ontology's representation method.

Acknowledgement

The Project is 2017 Annual Jiangxi Provincial Department of Education Science and Technology (GJJ161671), 2014 Jiangxi University of Traditional Chinese Medicine research research one of the results (2013ZR0073).

Reference

- [1] Yuan Lei, Zhang Hao, Chen Jing, Lu Jianfeng. Studies on the knowledge base construction mode of ontology-based knowledge model, *Computer Engineering and Applications*, 2006.30:25-27.
- [2] Deng Zhihong, Tang Shiwei, Zhang Ming, Yang Dongqing, Chen Jie. Research review on Ontology, *Acta Scientiarum Naturalium Universitatis Pekinensis (JCP Science Edition)*, 2002, (05):730-738.
- [3] Li Shanping, Yin Qiwei, Hu Yujie, Guo Ming, Fu Xiangjun. Research review on ontology, *Journal of Computer Research and Development*, 2004,(07):1041-1052.
- [4] Lu Ruling (Editor-in-chief). Knowledge engineering and knowledge science in the turn of the century, Tsinghua University Press, 2001.
- [5] Guus Schreiber et al. Knowledge Engineering and Management: The CommonKADS Methodology, China Machine Press, 2003.
- [6] Rao Yuan, Feng Boqin. Studies on ontology-based XML knowledge representation method, *Microelectronics Computer*, 2004; 21(9): 260-29.

- [7] Studer R, Benjamins V R, Fensel D. Knowledge Engineering, Principles and Methods. *Data and Knowledge Engineering*, 1998, 25(12): 161~197.
- [8] Gruber T. A Translation Approach to Portable Ontology Specifications, *Knowledge Acquisition*, 1993(5):199—220.
- [9] Zhang Jifang. Construction and application of ontology-based teaching domain knowledge base, *Time and Experience*, 20011, (4).
- [10] Perez A G, Benjamins V R. Overview of Knowledge Sharing and Reuse Components: Ontologies and Problem-Solving Methods in Proceedings of the IJCAI-99 Workshop on Ontologies and Problem-Solving Methods. Stockholm, Sweden, 1999:1~15.
- [11] Baader F, Horrocks I, Sattler U. Description logics as ontology languages for the semantic web. In: Hutter D, Stephan W. eds. *Festschrift in honor of Jorg Siekmann. Lecture Notes in Artificial Intelligence*. Springer, 2003.
- [12] Noy N F, McGuinness DL. *Ontology Development: a Guide to Creating Your First Ontology* [EB/OL] <http://protege.stanford.edu/publications/ontology-development/ontology101.pdf>, 2002.
- [13] Gruber T, 1995. Towards principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies* 43 (5/6): 907-928.
- [14] Li Jing, Meng Xianxue, Su Xiaolu. *Studies on construction methods and application of domain ontology*, Beijing: China Agricultural Science and Technology Press, 2009, s 7.