

HOSTED BY



ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: <http://www.elsevier.com/locate/jtcms>



Expression model for multiple relationships in the ontology of traditional Chinese medicine knowledge

Mingyue Gao^a, Lepeng Wang^b, Meng Cui^{c,*}

^a College of Basic Theory of Traditional Chinese Medicine, Beijing University of Chinese Medicine, Beijing 100029, China

^b College of Arts and Science, Beijing Union University, Beijing 100083, China

^c Institute of Information on Traditional Chinese Medicine, China Academy of Chinese Medical Sciences, Beijing, 10070 China

Received 1 July 2016; accepted 1 July 2016

Available online 5 July 2016

KEYWORDS

Multiple relationships;
Ontology;
Expression model;
In traditional Chinese medicine knowledge

Abstract *Objective:* To explore multiple relationships in traditional Chinese medicine (TCM) knowledge by comparing binary and multiple relationships during knowledge organization.

Methods: Characteristics of binary and multiple semantic relationships as well as their associations are described. A method to classify multiple relationships based on the involvement of time is proposed and theoretically validated using examples from the ancient TCM classic *Important Formulas Worth a Thousand Gold Pieces*. The classification includes parallel multiple relationships, restricted multiple relationships, multiple relationships that involve time, and multiple relationships that involve time restriction. Next, construction of multiple semantic relationships for TCM concepts in each classification using Protégé, an ontology editing tool is described.

Results: Protégé is superior to a binary relationship and less than ideal with multiple relationships during the constitution of concept relationships.

Conclusion: When applied in TCM, the semantic relationships constructed by Protégé are superior than those constructed by correlation and/or attribute relationships, but less ideal than those constructed by the human cognitive process.

© 2016 Beijing University of Chinese Medicine. Production and hosting by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

* Corresponding author. Tel.: +86 13910701699.

E-mail address: cm@mail.cintcm.ac.cn (M. Cui).

Peer review under responsibility of Beijing University of Chinese Medicine.

Introduction

Traditional Chinese medicine (TCM) entered the digital era when computer science was introduced into the research of the subject.¹ A large amount of ancient literature on TCM is stored in text. Software is used to present and summarize research results, and predict the evolution of the corresponding disciplines.^{2–6} As the expectation of using computers to simulate human thinking has increased, digitization not only means the simple storage of information, but also the storage of information in a way that simulates human thinking. The method for developing software has become object-oriented programming instead of structured programming. Data analysis has been transferred to big data techniques^{7,8} from previous statistical and data mining methods.^{9–11}

Big data analytics are the main method for researching huge, complicated and associated data, and can be applied to analyzing rules and forecasting related knowledge from data.¹² TCM data belong to intellect-intensive data and have the following features: integrity, hybridity, and relativity. Although the storage content of TCM information is not sufficiently big, Cui et al.¹³ believe that TCM data satisfy the requirement of full-scale data,¹⁴ which makes it possible to treat them as big data. Thus, the techniques for investigating big data can be used to process TCM data. Li et al.¹⁵ of the Chinese Academy of Engineering, hold that when analyzing big data, their underlying complex networks, that is, the relational networks that have knowledge structure, need to be analyzed. Thus, the prerequisite for analyzing information for a given field is a proper knowledge model, which can transform information stored in the form of natural language into a form that computers can understand, and make the simulation of human thinking possible.

At present, knowledge models for TCM have been proposed, possessing inquiry, search, and simple inference functions,^{16–25} and some models can even exchange information between systems.^{26,27} At the same time, the knowledge model for imitating the human brain is still being explored. The object-oriented development method considers the elements of the system as objects in the real world, using terms that are similar to those used in the given field, which makes a computer more similar to the human mind.²⁸ This is beneficial for analyzing, designing, and programming complex systems. Among these methods, the Unified Modeling Language (UML) is the standard modeling language that dominates the object-oriented field. Ontology is the main development tool for using UML to establish object-oriented models.²⁹ The semantic models developed by ontology simulate storing information, as the human brain does, which is a proper method for organizing and expressing knowledge, and demonstrates the relational network between pieces of information to a certain degree. The establishment of semantic networks developed by ontology is the precondition for investigating TCM knowledge using big data technologies.

The unified traditional Chinese medical language system (UTCMLS) is the most extensively used semantic network system in the TCM field. UTCMLS was built by the China Academy of Chinese Medical Sciences (CACMS) in 2002 using

ontology.¹⁶ Since then, many research institutions have begun to build semantic networks of TCM.^{18,21,30} The building of semantic networks simulates human thinking mainly through constructing relationships between concepts.³¹ For some time, this type of method has been limited to the binary relationship construct or has transformed multiple relationships into binary relationships.³² Semantic relationships in TCM are complicated and multiple relationships in TCM arise almost everywhere. Thus, the discussion and research into multiple relationships in semantic networks is still in its early exploration stage.

Information is the carrier of knowledge and knowledge reacts to information. Computers can store and understand knowledge only through information transformation. The characteristics of TCM information include three aspects. First is the integrality of the meaning. Every piece of information recorded in TCM texts is independent, including basic information on patients, such as age and sex; symptoms; syndrome pattern; and treatment prescriptions. A meaningful knowledge expression conveys a patient's essential state using a complete set of information as presented by the patient. Second is the hybridity in semantics and grammar. The smallest unit in TCM information is the concept. The semantic network construction of a computer is based on conceptual extraction. The concept of TCM information includes a verb, conjunction and preposition at the semantic level, and subject–predicate form, verb–object form and postpositive attributive form at the grammatical level. Third is the complexity of semantics. The same word or word group expresses a different concept and is a different part of speech in different language environments.

The aforementioned characteristics demonstrate that it is very difficult to build semantic networks based on TCM information. Solving the divergence between grammar and semantics is the basis of constructing a semantic network, followed by maintaining the completeness of text information determines the transformation and consistent degree of storing information by the human brain thought analogy.

Through an extensive investigation of ancient TCM works, we found that TCM knowledge is not only a binary relationship, but also a multi-relationship. Based on this understanding, we used the ontology software Protégé to construct a semantic network of the ancient TCM classic, *Important Formulas Worth a Thousand Gold Pieces (Beiji Qianjin Yaofang)*. Our was to develop a multiple relationship expression of TCM information to resolve the completeness of the meaning of TCM. *Important Formulas Worth a Thousand Gold Pieces* is a seminal work in Chinese medicine published in the Tang dynasty in 652. It was written by Sun Simiao, the physician of extraordinary renown in Chinese medical history. The book summarizes pre-Tang medical cases and is believed to have greatly influenced the development of TCM in China. The work lists about 5300 herbal formulas and discusses a wide range of medical topics including gynecology, pediatrics, internal medicine, external medicine, emergency treatment, food therapy and health preservation, acupuncture, among other subjects. We chose this book as our research focus because of its significance the development of TCM and its

Download English Version:

<https://daneshyari.com/en/article/1993133>

Download Persian Version:

<https://daneshyari.com/article/1993133>

[Daneshyari.com](https://daneshyari.com)