Ontology-Based Modeling of Clinical Reasoning in Traditional Chinese Medicine

Tong Yu, Jinghua Li, Hongjie Gao, Qi Yu, Meng Cui
Institute of Information on Traditional Chinese Medicine
China Academy of Chinese Medical Sciences
Beijing, China
E-mail: cui@mail.cintcm.ac.cn

Abstract—the integration of Traditional Chinese Medicine (TCM) and Western Medicine (WM) requires extensive interdisciplinary and cross-cultural collaboration. The productivity of integrated studies is often limited by the lack of understanding between practitioners from different cultural backgrounds. Information technologies, especially the Semantic Web and ontologies, can facilitate the exchange of medical information across discipline boundaries. We use Semantic Web technologies to engineer a TCM domain ontology named TCMOnto, covering categories such as basic theories, diagnostics, diseases, therapeutics, and medicinal treatments. TCMOnto formalizes the meaning of TCM concepts, and captures the semantic relations between them. This ontology can be used to represent knowledge resources such as medical rules, semantic queries, and semantic annotations for medical documents, and facilitate a variety of applications such as information integration and exchange, clinical intelligence, and knowledge management.

Keywords—traditional Chinese medicine; integrated medicine; ontology; semantic web; knowledge representation

I. INTRODUCTION

As an iconic cultural heritage of China, Traditional Chinese Medicine (TCM) is a comprehensive system of medicine characterized by its own theory and principles [1]. During the last five decades, there has been increasing interest in this cultural heritage in the community of Western Medicine (WM), leading to the endeavor to further develop TCM in the framework of Integrated Medicine (IM) [2]. The productivity of research, however, is often limited by the obstacles to exchange information and knowledge across culture and discipline boundaries. Notably, most WM researchers do not understand the theory and knowledge system of TCM. Information technologies can potentially bridge these cultural gaps.

In this paper, we introduce a community effort to investigate and codify the reasoning process underlying TCM clinical practices. We aim to transform this reasoning process from practitioners’ personal experience into a shared body of knowledge, in order to facilitate heritage preservation, clinical research, and education. An information technology named Ontology is adopted in this project. In computer science, ontology is defined as a formal, explicit specification of a shared conceptualization [3]. We assume that the TCM clinical reasoning is based on some unique form of clinical logic, which contains a set of concepts, facts, rules, and principles that could be formalized. We capture this logic by reviewing the literature in TCM domain, extracting the logic through text analysis, and represent the logic using a domain ontology.

TCM clinical processes are characterized by holism and treatment based on syndrome differentiation. TCM practitioners focus on patients rather than diseases. They examine the patient as an integrated system, and make holistic therapeutic strategies to promote health and enhance the quality of life, rather than focusing on specific diseases or symptoms. A clinical reasoning process is essentially focused on the syndrome of the patient, and contains two sub-processes: (1) syndrome differentiation, which takes the observed symptoms as the input, and generates the evaluation of the morbid condition as the output; and (2) syndrome intervention, which takes the diagnosed syndrome as the input, and generates the therapeutic plan as the output. These reasoning processes are recorded in a plurality of case records, which can be found in TCM classics, scholarly papers, and electronic medical records.

In this paper, we introduce the TCMOnto, a domain ontology that models the TCM clinical reasoning by representing the basic concepts, theories, rules and principles in formal languages. We engineer the TCMOnto ontology with the following process: (1) conduct a systematic analysis of literature and extract a set of concise statements about the clinical logic; (2) codify the meaning of TCM concepts, and define the semantic relationships between them; (3) represent domain knowledge as formal statements and rules. The resulting ontology covers the major areas of TCM domain, including syndrome classification, therapeutics, medicinal herbs, and formula study.

We develop a knowledge management system to store and manage this ontology. Automated reasoning methods can then be used to infer consequences of the ontology. This ontology is accessible via a Web service, which enables logic-based query-answering. This ontology can facilitate a variety of applications such as information integration, information retrieval, clinical intelligence, knowledge reference, knowledge management, knowledge discovery and knowledge sharing.
In the rest of the paper, we will discuss the process and methods of ontology engineering, and also the content and applications of this ontology.

II. PROCESS AND METHODS OF ONTOLOGY ENGINEERING

A. Semantic Web Technology

We implement this ontology using the Semantic Web [4], which is an emerging Web technology that allows data to be shared and reused across application, enterprise, and community boundaries. The Semantic Web technology provides standard languages, such as Resource Description Framework (RDF) [5], RDF Schema (RDFS) [6], and Web Ontology Language (OWL) [7], for knowledge representation, data integration, and automatic reasoning. The Semantic Web has two aspects: the Semantic aspect and the Web aspect:

1) The Semantic aspect: the term “Semantic” refers to an expression’s meaning, or the things that the expression refers to. We say that a machine “understands” an expression, if it is aware of the things behind the expression and grasps their relationships. The Semantic Web provides formal languages to express information with explicit semantics, which any intelligent agents (as well as humans) can “understand”.

2) The Web aspect: Semantic Web languages are Web-compatible, which enables ontologies to be published on the Web and linked to relevant ontologies and data, in order to facilitate interdisciplinary collaboration on a global basis.

The Semantic Web enables ontologies to be reused and connected to existing ontologies, and also supports Web-based collaboration in ontology engineering between multiple medical practitioners. Therefore, the Semantic Web provides a sound technical solution for ontology engineering.

B. Ontology Design

TCMonto contains two major parts: a TBox (or terminology box) that provides a shared vocabulary that standardizes the naming, definition, and categorization of the TCM concepts that are used to express the TCM clinical logic; and An ABox (or assertion box) that contains logical statements and rules that formalizes the clinical logic.

The ABox is implemented in Resource Description Framework (RDF) [5], which is a standard model for data interchange on the Web. TBox is implemented using the vocabulary of RDFS [6] and OWL [7], and is expressed using the Manchester syntax [8].

The major purpose of TBox is to group things into classes, which form a class hierarchy. We can determine the membership of the class in two ways: (1) declaring membership explicitly; or (2) inferring membership based on facts about an individual. In addition to class definitions, we also organize properties into a hierarchy. The ontology can be characterized by two graphs: (1) the class subsumption graph is a directed graph that represents subsumption relationships among classes; and (2) the property graph represents the semantic relationships between classes.

We reuse some well-known and widely-adopted vocabularies. For example, we reuse the Unified Medical Language System (UMLS) [9], a repository of biomedical vocabularies developed by the US National Library of Medicine, by (1) defining general TCM concepts as subclass of UMLS Semantic Types, and (2) reusing UMLS Semantic Relations to express relationships between TCM concepts. We also reuse the terms from Dublin Core [10] and SKOS [11] when appropriate.

C. Ontology Engineering Process

In this project, we first conduct a systematic text analysis of TCM literature in order to extract a set of concise statements about the TCM clinical logic. Next, we construct a formal terminology that codifies the meaning of complex TCM concepts. Finally, we construct semantic graphs that represent semantic relationships between these concepts. The engineering of TCMonto takes the following procedure:

1. Determine the TCM knowledge assets to be codified.
2. Conduct a systematic review of knowledge assets.
3. Generate a set of concise statements that express the TCM clinical logic.
4. Identify key concepts, clarify their meaning, and generate a description for each concept.
5. Formalize the semantic relationships between concepts by constructing the class hierarchy and the association network.
6. Express clinical facts as a set of binary relationships between concepts.
7. Check and revise the ontology content.
8. Publish the ontology via a Web-based ontology service.

D. Ontology Engineering Platform

We have built a Web-based ontology engineering platform to facilitate the collaborative engineering of TCMonto, which allows knowledge engineers to edit and explore ontology online. Users can use this platform anywhere with their Web browsers, and don’t need to install any special client program. The ontology is stored and managed in a back-end knowledge base, which is also accessible to client programs via a Web-based programmable interface.

III. DETAILED DESCRIPTION OF TCMONTO CONTENT

The TCMonto ontology formalizes the basic concepts, theories, rules, and principles in clinical reasoning, such as syndrome classification and differentiation, therapeutic principles, and herbal formula combination. It consists of the following components: (1) classes (such as Drug, Herb, and Disease) that define types of things, (2) properties (such as treats, consists of, and interacts with) that defines binary relationships between things, (3) individuals (such as qi, yin, and wood) that define things that exist in the domain, which are associated with each other through properties, and which are grouped into classes, and (4) facts and rules that capture the knowledge content of TCM domain. The main components of TCMonto are as follows: TCM Syndrome Classification, TCM Therapeutics, TCM Medicinal Herbs, and TCM Formula Study. A detailed description of TCMonto content is presented as follows.
A. TCM Syndrome Classification

The “TCM Syndrome Classification” module represents the classification system of TCM syndromes, the relationships between syndromes and symptoms, and the logic of syndrome differentiation.

TCM practitioners typically characterize patients with TCM Syndromes, which are integrated patterns of human morbidity indicating the hidden cause of observable symptoms. A syndrome type is a common syndrome mode with a standard name. Each syndrome type is defined as a composite concept in terms of yin-yang, five phases, viscera and bowels, etc. We model TCM syndrome types as the Syndrome Hierarchy, a classification system that classifies syndromes into different classes.

The Syndrome Hierarchy is organized according to the Eight Principles of Syndrome Differentiation. According to the Yellow Emperor’s Inner Canon, there were eight basic methods of syndrome differentiation: Yin, Yang, Exterior, Interior, Cold, Heat, Deficiency (Xu) and Excess (Shi). The Eight Principles are used as the eight top-level classes of the Syndrome Hierarchy.

Here, we focus on the analysis of Deficiency (Xu) and Excess (Shi), which are two principles that help analyze and generalize the opposing forces between the body resistance and pathogenic factors during the course of a disease. Owing to various disease preventing factors inside the body, Deficiency can be further classified into Qi Deficiency, Blood Deficiency, Yang Deficiency, Yin Deficiency and Body Fluid deficiency. For example, the syndrome Yang Deficiency indicates that it is the Yang that is deficient; or it is the Yang in the body that is affected by Yang Deficiency. We define the class Yang_Deficiency as follows:

- Deficiency that affects the Yang.

We can also specify which organ is in an improper state. For example, we say that Kidney Deficiency is the Deficiency that affects the Healthy Qi (can be Qi, Blood, Yin or Yang) in Kidney, and define the class Kidney_Deficiency as:

- Deficiency that has_location the Kidney.

In particular, the syndrome Kidney Yang Deficiency indicates that the Yang in the Kidney is deficient. Here, the term Kidney Yang (also called primordial yang, true yang or true fire) is a metaphor that represents the Yang in the Kidney, which is the foundation of the Yang Qi of the whole body, and warms and promotes the functions of the organs and tissues. We can represent the definition for the class Kidney_Yang_Deficiency as follows:

- Deficiency that has_location the Kidney and affects the Yang.

In addition, syndrome types can be combined into more complex syndrome type. For example, Spleen Kidney Yang Deficiency is a “pathological change characterized by insufficient yang qi of the spleen and kidney with endogenous cold and flooding of water”, and can be defined as a combination of Spleen Yang Deficiency and Kidney Yang Deficiency. Accordingly, the class Spleen_Kidney_Yang_Deficiency can be defined as follows:

- Spleen_Yang_Deficiency and Kidney_Yang_Deficiency.

Based on these ontological definitions, a reasoner can infer the transitive hierarchical relations between syndromes, which form the Syndrome Hierarchy.

B. TCM Therapeutics

The “TCM Therapeutics” module represents TCM therapeutic principals and methods, and their applications in therapeutic practices.

TCM practitioners treat the human body as an organic whole, and aim to address the identified syndrome in treatment in order to restore the balance and harmony of the human body. The therapeutic principles are mainly derived from the Eight Principles of syndrome differentiation. For example, when an Excess syndrome is identified, the treatment focuses on reducing the excess and eliminating the factors that brought about the excess in the first place. If a Deficiency is identified, the treatment focuses on tonifying (means nourishing and invigorating) the deficiency to ensure that energy is maintained at a healthy level. Similarly, Heat should be cooled; Cold should be warmed; Dampness should be expelled; and Phlegm should be resolved. Accordingly, we define the therapeutic method Tonification and as follows:

- Therapeutic_Method that tonifies some Thing.

Similarly, we define the therapeutic method Purgation as follows:

- Therapeutic_Method that purges some Thing.

Also, we model two basic therapeutic principles with formal rules:

1. Treat Deficiency by Tonification (treat deficiency conditions by supplementation or tonification) is model by the following rule:

   (affected by some Deficiency) implies (treated by some Tonification).

2. Treat Excess by Purgation (treat excess conditions by purgation or reduction) is modeled by the following rule:

   (affected by some Excess) implies (treated by some Purgation).

According to the above principles, the major strategy to treat Kidney Yang Deficiency, is to tonify Kidney Yang; whereas the major strategy to the Kidney Yang Excess, is to reduce Kidney Yang. Various therapies, such as medicinal herbs, acupuncture and moxibustion, massage, etc., can be used to implement these therapeutic strategies. However, we only focus on TCM medicinal herbs and TCM formulae in the engineering TCMMonto ontology.
**C. TCM Medicinal Herbs**

The “TCM Medicinal Herbs” module represents the knowledge of medicinal herbs. TCM is characterized by the use of medicinal herbs as therapies. We define an *Herb* as a plant used as herbal material to treat some syndromes. The properties used to characterize the *Herb* include: efficacy, safety, toxicity, dosage (form, amount, etc.), allergies, indication, flavor, meridian entry gui jing, etc. Accordingly, we define the class *Herb* as follows:

- Drug that uses some Plant
  - and has _meridian_entry_ some TCM_Organ
  - and has _flavor_ some Herbal_Flavor
  - and has _nature_ some Herbal_Nature.

Here, we use the *Asian Ginseng* (simply Ginseng, or Ren Shen), a well-known herb used in TCM, as an example of representing herbs. *Ginseng* has the taste of **Sweet** and **Bitter**, the quality of **Warm**, and the efficacy of **tonifying the healthy qi** (such as **qi**, **blood**, and **yang**) of the human body. Therefore, *Ginseng* can tonify yang when it is deficient, in order to restore the balance of the body. Accordingly, we define the ontological definition of *Ginseng* as follows:

- Herb that treats the **Qi_Deficiency**
  - and has _flavor_ the {**Sweet**, **Bitter**}
  - and has _nature_ the {**Warmth**, **Neutral**}
  - and tonifies the {**Yang**, **Qi**, **Blood**}
  - and fortifies the Spleen
  - and replenishes the Lung.

The physician constructs a logical justification for a decision of intervention based on therapeutic principles. For example, a physician identify that a patient has the syndrome *Yang Deficiency*, which is caused by the deficiency of the *Yang*. The physician reasons that an intervention is needed to tonify the deficient yang in order to restore the proper condition. In addition, the physician recognizes that the herb *Ginseng* has the efficacy to tonify the *Yang*, and therefore use *Ginseng* as a therapy. This reasoning process can be formalized as the following process:

1. Represent the therapeutic principle “Treat Deficiency by Tonification” as a therapeutic rule:
   - (affected by some Deficiency) implies (treated by some Tonification)
2. Identify the fact about the patient:
   - Yang affected by Yang Deficiency.
3. Match the fact with the rule, resulting into the assertion that “yang should be treated by a tonification”:
   - Yang treated by [a Tonification] .
4. Given the fact that:
   - Ginseng is a Tonification,
   - a therapy that uses Ginseng can be prescribed to treat the patient.

**D. TCM Formula**

The “TCM Formula Study” module represents combination of medicinal ingredients, composition of prescriptions and the clinical uses of herbal medicines, etc.

Traditional Chinese herbal therapies can be characterized by the use of multi-herb formulae. In history, TCM pharmacists primarily compounded therapies as mixtures of multiple herbs, and establish an elaborate system of herbal formulae. A TCM Herbal Formula can be defined as a prescription that consists of a set of herbs, which interact with each other in order to achieve certain therapeutic functions. A formula typically undergoes certain processing method, such as decoction, in order to enact the proper interaction between herbs.

Therefore, the specification of a formula includes the enumeration of its individual herbs, their dosage and functional role, the rationale of herb combination, the mechanism of action, and its relationships with other medical concepts. TCM pharmacists adopt a social metaphor of formula combination, and interpret the formulae system as a “social network” of herbs. They have the essential principle that a formula should embody a proper herb-community involving hierarchical social relationships, between a single dominant figure, the *Sovereign Herb*, and a set of subordinate figures such as *Minister Herbs*, *Assistant Herbs*, and *Courier Herbs*. A formula can be interpreted as a collaboration of its member herbs which play different roles.

A well-known example of this social metaphor is the ancient herbal formula Four-Gentlemen Decoction (FGD for short), which contains four “gentlemen” who interact with each other to achieve the efficacy of “tonifying the qi”. The four herbs are: (1) *Ginseng* (*Gins* for short), the *Sovereign Herb* that directly tonifies the qi; (2) *Atractylodis* (*Atr* for short), the *Minister Herb* that aids the tonification of the qi by nurturing the spleen and drying the dampness of body, (3) *Scelrotium* (*Scler* for short), the *Assistant Herb* that assists the king and the minister in strengthening the spleen, and (4) *Glycyrrhizae* (*Glyc* for short), the *Courier Herb* that warms and harmonizes the spleen (so as to facilitate the absorption of other herbs).

We represent the formula combination as follows: (1) define FGD as Formula and *Gins*, *Atr*, *Scler*, *Glyc* as Herb, (2) assert the containment relationships between a formula and herbs, and (3) annotate the interactions between herbs. Accordingly, we define the ontological definition of *Four_Gentlemen_Decocion* as follows:

- Formula that
  - has _preparation_form_ the Decoction
  - and sovereign the Ginseng
  - and minister the Atractylodis
  - and assistant the Sclerotium
  - and courier the Glycyrrhizae
  - and treats the Spleen_Qi_Deficiency
  - and tonifies the Qi
  - and fortifies the Spleen.

---

136
IV. APPLICATION

TCMonto can be used to generate semantic annotations for information resources in TCM domain. Most of information resources, such as TCM classics, scholarly papers, and images, are unstructured, which causes the difficulty of sharing and retrieval of information resources. For example, a scholarly paper titled “Huperzine A, A Potential Therapeutic Agent for Treatment of Alzheimer’s Disease” [12] includes the following unstructured text: “...HupA was isolated from Huperzia serrata (Thunb) Trev. by Liu and his coworker in the early 1980’s...”. It would be useful yet difficult for a software agent (such as a search engine) to recognize the fact that this paper asserts that “Huperzia Serrata contains Huperzine A”.

A solution is to annotate unstructured resources in terms of a shared domain ontology, which make it easier for agents to process them. Here, semantic annotations for a document are RDF graphs that describe the semantic content of the document in terms of a domain ontology; An example of semantic annotation is as follows:

```
{paper1 dc:subject {Huperzia_Serrata contains Huperzine_A.}},
```

which means that a subject of paper1 is that “Huperzia Serrata contains Huperzine A”.

We have constructed a Web-based platform for TCM practitioners to generate semantic annotations for TCM literature in terms of the TCMonto. Semantic content can be extracted from documents through both automatic and manual annotating processes. These semantic annotations were collected and managed in a knowledge base, which serves as the semantic metadata for Semantic Web applications in TCM domain.

V. CONCLUSIONS

A shared conceptualization of TCM clinical logic is fundamental to the preservation, investigation, utilization, and further development of this cultural heritage. The characteristics of TCM clinical logic, such as ancientness, fuzziness, diversity, and complexity, require advanced knowledge modeling techniques. To deal with these issues, we model the TCM clinical logic as a domain ontology, and use the Semantic Web to implement the ontology. The resulting domain ontology, namely TCMonto, is a standard codification of the TCM clinical logic. TCMonto can foster the understanding of TCM clinical reasoning, and facilitate the communication, collaboration, and decision-making in clinical studies of integrated medicine.

ACKNOWLEDGMENT

This work is supported by “the Fundamental Research Funds for the Central public welfare research institutes (NO. ZZ080324, NO. ZZ090305, NO. ZZ070804, NO. ZZ090301”).

REFERENCES