Toward Global Health Interoperability: Ontology Development of Traditional Chinese Medicine Through Comparative Analyses

Au Vo
Claremont Graduate University, USA
au.vo@cgu.edu

Kun Liu
Claremont Graduate University, USA
kun.liu@cgu.edu

Thomas Horan
Claremont Graduate University, USA
tom.horan@cgu.edu

Yusheng He
Peking University, Beijing, China
heyusheng263.net

Abstract

The World Health Organization’s “Health for One” initiative calls for an interoperability between various medical practices around the world. As one of the oldest medical practices, Traditional Chinese Medicine (TCM) has been a major global area of practice and research, and is ripe for enhanced interoperability. Such interoperability relies on a strong ontological approach to create a reusable and shareable knowledge base. This ontology enables TCM to be used in conjunction with the Health Information Technology. To enable universal interoperability, this paper proposes a framework to examine current applications of TCM ontology. The framework integrates the evaluation of two versions of TCM ontology: the current China SNOMED-CD based implementation and TCM’s international interpretation (International Classification of Traditional Medicine). The framework also suggests a creation of a new ontology that is compatible with international standards and Health Information Technology standards.

1. Introduction

Ontology is a widely adopted methodology for capturing and disseminating knowledge within the medical domain. The usage of ontology has been well-documented in variety of medical fields that include biomedical, medical knowledge, and disease domains.

Although ontology is ubiquitous to the domain, different goals lead to different ontologies. As a result, the implementation of ontology has created highly fragmented ontological structures within the medical domain. To address variety of needs, organizations and countries have started adopting medical ontologies that could be extended and used universally. Two ontologies have emerged and established themselves as international standards: International Classification of Diseases (ICD) and Systematized Nomenclature of Medicine - Clinical Terms (SNOMED-CT).

International organizations, such as the World Health Organization (WHO), often support development of medical practice interoperability standards among countries. The WHO’s initiative “Health for One” promotes healthy living and comprehensive care. The WHO has also starting to survey indigenous medical practices as an alternative to conventional medical practices. One of the practices in focus is the Traditional Chinese Medicine (TCM).

Originated 5,000 years ago, TCM is projected to be one catalyst for the “Health for One” initiative [13]. However, to enable TCM to become interoperable with western medical practices, TCM knowledge has to be encoded in such a way that it could be used universally. Attempts have already been made and TCM ontology reflecting SNOMED-CT ontological approach has been applied. This enables TCM to be used with Electronic Health Records (EHR) in China [2]. This ontology is focused exclusively on EHR implementation. Another ontology to address other TCM needs is being formulated. This ontology uses the framework defined as the part of ICD-11. This ontology attempts to make TCM a part of international interoperability standards.

Neither one of these ontology standards meets all the needs of various processes where TCM clinical data is used. In this paper, we propose a framework for integrating the two ontology standards when utilizing TCM clinical data. The framework can be used for universal TCM implementation and adapted for other indigenous medical practices such as Ayurvedic Medicine and Unani Medicine.

In section 2 of this paper, we give a brief background of TCM. In section 3, we review ontology usage in the medical domain. Section 4 provides the framework. Section 5 examines SNOMED-CT and
ICD-11 ontological approaches for TCM and provides preliminary insights on their similarities and differences. The final section discusses the findings.

2. Background

Traditional Chinese Medicine (TCM) is a 5,000-year-old medical practice. Since its inception, TCM is an essential part of the Chinese culture and health care. TCM is a vital medical practice in modern China. In China, as many as 60% of the Chinese population prefer TCM consultation at least once a year, while as many as 75% of the Eastern Asian population use TCM at least once a year [3].

TCM practice is a part of other cultures as well. Thus, not only is TCM prevalent in Japan, Korea, and Vietnam, even many western countries have started to use TCM as an alternative and/or complementary treatment. In 2010, in the USA, $7.6 billion were spent on the TCM products [3].

The WHO identified TCM as a major part of traditional medicine. WHO defines traditional medicine as the sum total of knowledge, skills and practices based on the theories, beliefs and experiences indigenous to different cultures that are used to maintain health, as well as to prevent, diagnose, improve or treat physical and mental illnesses [29].

TCM is different from western medical practice. TCM possesses a plethora of unique theories, methods and diagnoses. TCM not only identifies and treats illness and prevents disease but also optimizes wellbeing and wellness sustainability. Chinese perceive both TCM and western medical practice as complementary to each other. They prefer to have access to both medical practices. Recognizing the importance of including both TCM and the western medical practice in Chinese medical practices, the Chinese Ministry of Health established the China Academy of Chinese Medical Sciences (CACMS) in 1955 to help preserve, expand, and integrate TCM with the western medical practices.

To meet the need for preservation and adoption of TCM in the Health Information Technology era, CACMS has collected and organized TCM terminologies. CACMS uses the SNOMED-CT ontology to create the knowledge base for TCM. The use of this ontology is relevant because using this ontology, Information Technology can be applied to support TCM practices.

The motivation for this study derives from CACMS’s interest of developing a TCM ontology in order to guide their efforts in EHR implementation and in the manner that is consistent with international standards, i.e. ICTM, and interoperable with Health Information Technology. As a result, a collaboration between CACMS and the research team has been formed to conduct a multi-phase study. This comparative analysis presented in the paper is the first phase.

3. Overview of Ontology in Medical Domain

A comprehensive review of ontology was conducted focusing on specific use within the medical domain. This section presents a brief overview of ontology method, followed by a brief discussion about different types of ontologies applied in medical domain. We further classify each exemplar ontology into their appropriate types.

3.1. Ontology General Description

Ontology has been a preferred method for knowledge management research. Initially, the notion of ontology was developed in an effort to help develop knowledge management principles for deploying semantic web [6]. The goal of ontology is to classify the most complex and widespread knowledge systematically. Ontology was used to help with the representation, organization, acquisition, creation, usage, and evolution of knowledge [14]. With the advent of Information Systems, ontology-driven classification has been the most utilized principle in knowledge management [23].

Ontology is a kind of taxonomy with structure and specific types of relationships between terms [24]. In any particular domain, ontology serves as a declarative model that can be used to present different constructs, their attributes, and the interrelationships between the constructs. There are four major types of ontologies: general ontology, domain ontology, application ontology, and reference ontology [1]. General ontology is an overarching knowledge representation, notwithstanding the subsequent task-related level. Domain ontology pertains to specific domain of interest. Application ontology, as its name suggests, is an applied ontology that is pertinent to a specific task. Finally, reference ontology acts as a reference knowledge base that can be shared among different
domains [1]. For this research, we are specifically interested in the domain ontology viz., the ontology in the medical domain.

### 3.2. Medical Domain Ontology

Medical domain ontology consists of several ontological approaches to classify and depict medical information in the most useful manner. The most notable ontology subclasses are biomedical ontology, disease ontology, ontology of medical terminology, ontology of medical knowledge, and ontology for clinical findings. The majority of the existing ontologies can be classified into one or more of these medical ontology types. The table (Table 1) below gives examples of ontology in each subclass:

<table>
<thead>
<tr>
<th>Name</th>
<th>Main Medical Domain</th>
<th>Ontology Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>OpenCyc</td>
<td>N/A</td>
<td>General</td>
</tr>
<tr>
<td>WorldNet</td>
<td>N/A</td>
<td>General</td>
</tr>
<tr>
<td>GALEN</td>
<td>Biomedical</td>
<td>Domain</td>
</tr>
<tr>
<td>UMLS</td>
<td>Biomedical</td>
<td>Domain</td>
</tr>
<tr>
<td>LinkBase</td>
<td>Medical Knowledge</td>
<td>Domain</td>
</tr>
<tr>
<td>ICD</td>
<td>Medical terminology</td>
<td>Domain</td>
</tr>
<tr>
<td>SNOMED-CT</td>
<td>Medical terminology</td>
<td>Domain</td>
</tr>
<tr>
<td>FMA</td>
<td>Biomedical</td>
<td>Reference</td>
</tr>
<tr>
<td>Neurological</td>
<td>Disease</td>
<td>Application</td>
</tr>
<tr>
<td>Disease Ontology</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 1: Medical Ontology Exemplars**

OpenCyc and WorldNet are the overarching ontologies that are used for classifying common knowledge and natural language processing. The goal of these ontologies is to determine constructs at an abstract level to be used in any field. For instance, Animal Physiology can use OpenCyc to classify its constructs, while cognitive psychology utilizes WorldNet [7]. Generalized Architecture for Languages, Encyclopedias, and Nomenclatures in medicine, GALEN, is an ontology that represents a wide variety of medical concepts. GALEN can be used to identify medical procedures [25]. Unified Medical Language System, UMLS, contains more than one million medical constructs. UMLS is intended to be used by health care professionals by providing a medical encyclopedia [1]. LinkBase is a comprehensive ontology that covers procedures, anatomy, pharmaceuticals disorders, and anomalies. LinkBase specializes in transforming free text into medical information [28]. Foundation Model of Anatomy (FMA) is a reference ontology regarding the body constituents. As an extension of UMLS, Foundation Model of Anatomy allows alignment between other ontologies with its own structure [22]. Neurological Disease Ontology, as the name suggested, focused on neurological disease treatment and study [5].

ICD and SNOMED-CT [12] are the most common medical terminology ontologies in the medical domain. The ICD and SNOMED-CT are used extensively in clinical settings. ICD-10, the current version of ICD, has been adopted in more than 100 countries. Similarly, SNOMED-CT is a critical component of EHR. EHR has become a catalyst for adoption of the Health Care Information Technology.

As the medical domain evolves, so does its knowledge base. Thus, ontologies have to be constantly updated to capture the constructs effectively. For example, there has been exploration on the extended use of ontology for the development of next version of ICD viz., ICD-11 [26] [27]. Efforts to combine SNOMED-CT and ICD-11 have also been explored in using ontology-based system [4]. The collaborative ontology creation has been a prominent topic in the Stanford’s Biomedical Informatics [18].

The ICD-11 and SNOMED-CT based ontologies do not address all the needs of medical domain by themselves. Information Technology applications to support TCM may need an ontology that is based on a combination of these two standards.

### 4. Research Methodology

The aim of this research is to formulate a framework that combines the two ontology standards used in medical domain to meet the needs of TCM. In this analysis (as shown in Figure 1). TCM’s SNOMED-CT ontological approach is referred as CACMS and TCM’s ICD ontological approach as ICTM. ICTM is a top-down approach. ICTM is based on an overall framework proposed by a committee-based set of experts. The ICTM principles guide the creating of classes, subclasses, and terminologies. These terminologies are further organized in appropriate classes. CACMS is a more of an inductive approach, where the existing TCM terminologies are collected, sorted, and classified into ontology.
As shown in the Figure 1, the research framework identified the attributes of these systems, and the first stage of our analysis to compare the structures as well as create ontological artifacts. Based on findings from this first stage, and expanded research model will be created to validate the findings through use cases. Finally, integration of ICTM and CACMS depends on actual usage of the medical domain.

In our data collection phase, we gathered the ontological data from both WHO and CACMS. This data is of a qualitative sort. In this first phase, qualitative data is appropriate for our comparative analysis. For ICTM data, we found the tentative ICTM ontology alongside with the ICD-11 beta draft. For CACMS data, we have extracted the ontological structure in CACMS live databases.

Comparative analysis was done to examine the similarities and differences of the two approaches. A key part of the comparison was to analyze the nature, structure, and scope of TCM within each domain. This analysis informed the future ontology development and validated its utility. In the section below, we detail our findings with regard to TCM and ICTM. We also give a comparative analysis of the two methods.

The comparative analysis was conducted through the use of Protégé, an ontology application tool. Using Protégé, CACMS and ICTM ontologies were coded and displayed for comparison. Protégé specializes in displaying and manipulating ontology for knowledge discovery and knowledge dissemination. Gennari discussed the evolution of Protégé and the impact of Protégé on the knowledge-based systems [8]. Noy et. al. [20] discussed how Protégé can be used for knowledge acquisition. Nyulas [21] and Horridge [11] have explored the multilingual challenges of using ontology to classify traditional medicine for the WHO.

5. Research Findings

5.1. International Classification of Traditional Medicine

ICD is an ontology for medical domain knowledge championed by the WHO. Its current version, ICD-10, was inducted in 1992. Currently, ICD-10 has been adopted by more than 100 countries. Some countries applied ICD-10 without any modification, while others modified ICD-10 to fit their current information infrastructure.

In an effort to provide updated and universal medical classification, the WHO has started to work on a newer version of ICD viz., ICD-11. Currently ICD-11 is in the beta phase and is scheduled to be completed in 2017. The WHO proposal answered the call for universal health care from the community by creating a separate chapter within ICD-11. This chapter called International Classification of Traditional Medicine (ICTM) captures the diseases and related health problems that are diagnosed in using indigenous medical practice around the world. ICTM was scoped to capture eastern medicine, namely indigenous medical practices that are prevalent in China, Korea, and Japan. From the overall guidelines and the existing ICD-11 ontological structure, experts develop ICTM ontological structure. ICTM focuses extensively on disease reporting that currently consists of more than 400 TCM medical terminologies and is expected to grow exponentially.
5.2. China Academy of Chinese Medicine Sciences

In contrast to ICTM, CACMS creation was based on a bottom-up approach. As a designated institution for persevering TCM, CACMS has collected virtually all TCM medical terminologies. CACMS, after examining these terminologies and their clinical practices, has come up with an ontology based on SNOMED-CT. Mao and Yin [16] created an ‘upper’ level of the ontology for TCM development. Zhou et.al. [30] created an ontological approach to unify TCM terminology knowledge. Further development showed the formation of a data warehouse for TCM knowledge discovery [31].

To amalgamate and standardize TCM clinical practices, CACMS used a framework (Figure 2) to gather TCM knowledge.

![Figure 2: TCM Knowledge Framework](image)

The three-process framework is used for enabling TCM medical terminologies for clinical practice. First, an exhaustive list of TCM terminologies was gathered. The consolidation of these terms was classified using a predefined epistemology. Afterwards, the terminologies were grouped under particular clinical terms set that could be used directly in a clinical practice. In the end, the clinical terminology sets provide the basis for different clinical templates. These templates are EHR-compliant.

5.2.1. Classifying TCM Clinical Terminologies.

TCM clinical terminologies were classified using SNOMED-CT standards. SNOMED-CT has been regarded as a standard reference that specializes in recording medical records accurately [15]. Thus, to enable TCM in clinical practices and integrate with the current clinical recording infrastructure, CACMS has modified SNOMED-CT to fit with TCM requirements. Using this modification, TCM terminologies were sorted and classified into their respective classes.

This bottom-up approach makes use of the already existing medical terminologies in TCM. Using SNOMED-CT foundation classes namely body structure, clinical finding, linkage concept, special concept, quantifier, procedure, and pharmaceutical, TCM experts attempted to categorize each terminologies. Classes such as prescription, disease and syndromes, theory, experiment, and context, were created to accommodate the nuances of TCM terminologies. Overall, the classification of more than 350,000 terminologies was divided into 12 categories.

![Table: 12 Categories of TCM](image)

<table>
<thead>
<tr>
<th>Category</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical Finding</td>
<td>Prescription</td>
</tr>
<tr>
<td>Disease and Syndrome</td>
<td>Body Structure</td>
</tr>
<tr>
<td>Theory</td>
<td>Context</td>
</tr>
<tr>
<td>Experience</td>
<td>Quantifier</td>
</tr>
<tr>
<td>Procedure/Treatment</td>
<td>Linkage Concept</td>
</tr>
<tr>
<td>Pharmaceuticals/apparatus</td>
<td>Special Concept</td>
</tr>
<tr>
<td>Prescription</td>
<td></td>
</tr>
<tr>
<td>Body Structure</td>
<td></td>
</tr>
<tr>
<td>Context</td>
<td></td>
</tr>
<tr>
<td>Quantifier</td>
<td></td>
</tr>
<tr>
<td>Linkage Concept</td>
<td></td>
</tr>
<tr>
<td>Special Concept</td>
<td></td>
</tr>
</tbody>
</table>

5.2.2. Forming Clinical Terminology Set.

The justification for creating the TCM medical terminologies repository using a modification version of SNOMED-CT is to enable the nation’s transformation into Health Information Technology. For TCM terminologies to be a part of China health, it is important that TCM terminologies are EHR compliant.

TCM experts group the terminologies into different sets, according to their needs. Based on various recording procedures in different institutions within CACMS, different terminology sets were constructed. While the sets make use of the relevant terminologies, each set pertains to a specific medical practice. The propagation of TCM terminologies set occurred organically.

5.2.3. Creating TCM Clinical Templates.

The creation of the terminology sets is a prelude to medical practice adoption. The sets would then be adopted by each CACMS institution. To enable the terminology sets in clinical settings, clinical templates are formed.

5.3. Protégé analysis between CACMS and ICTM
CACMS and ITCM has started to formulate their ontologies with different approaches and different purposes. Even though they are rooted within TCM domain, they partially resemble each other. This is consistent with a notion that different purpose will lead to different ontology [19].

The comparison between ontologies reflects a similarity in the thought process at the highest level of ontology. As we traverse down in the lower levels, the differences become apparent. We illustrate these finding by examples in the next two sections.

### 5.3.1 Comparability between CACMS’s Disease and ICTM’s Disorder

CACMS and ICTM employ different naming convention for their classes. At the top level, while “Disease” and “Syndrome” are used in CACMS, “Disorders” and “Patterns” are used in ICTM. ICTM has another class called “Signs and Symptoms”. Speculations regarding the correlation regarding this class with CACMS class have not been made. This is due to insufficient information in the class. CACMS’s “Disease” class has seven subclasses, in contrast with 15 subclasses found under ICTM’s “Disorder”. At this level, several classes are analogous. For instance, “Ear, Nose, and Throat” is a subclass in both ontologies, with a small variance in naming convention. “Gynecological Disease”, “Orthopedics Disease”, and “Pediatrics Disease” in CACMS are comparable with “Female Reproductive System Conditions Disorders”, “Bone, Joint, and Muscle System Disorders”, and “Childhood and Adolescence Associated Disorders” in ICTM, respectively. Most interestingly, CACMS combines internal-organ classification into a subclass called “Internal Medicine”, while ICTM enumerates these subclasses at the same level. These are depicted in Figure 4.

### Figure 4 - Top levels of CACMS (left) and ICTM ontologies (Right)

CACMS and ICTM diverge dramatically at the lower levels. Though the subclasses look comparable, they exhibit differences in terminologies used. For instance, in ICTM’s “Zhang-Fu Symptom Complex”
and ICTM’s “Organ System Patterns”, each subclass utilizes internal organs to classify terminologies (Figure 5). While ICTM specifies the distinction between different organs, i.e. heart, kidney, liver, lung, and spleen, CACMS views internal organs in more granularity and combine them together. In “Zhang-Fu Symptom Complex” class, there are subclasses with internal organs combinations such as heart-small intestine, kidney-bladder, lung-large intestine, and spleen-stomach.

These discrepancies augment with the inclusion of the “Zhang-Fu Accompanying Syndrome”. In this subclass, CACMS combines between the internal organs and different TCM theories such as qi, yin, and yang to enumerate the terminologies. The combination creates a high number of terminologies that are closely related. Thus, it is difficult to determine the correlation of a specific term. For instance, given the ICTM’s “Liver Qi Deficiency Pattern”, it is improbable to determine whether this terminology corresponds to any of the following CACMS’s terminologies: “Heart-Kidney Qi Deficiency Syndrome”, “Qi Deficiency and Blood Static of Liver-Stomach Syndrome”, “Liver-Spleen Qi and Blood Deficiency Syndrome”, and “Liver-Spleen Qi and Yin Deficiency Syndrome.

5.4. CACMS Clinical Practices as a Validation Metric

In addition to international interoperability and Health Information Technology compliant, TCM ontology would have to be consistent with current TCM medical practices. In Figure 6, we depict the generalization of TCM medical practices. The practices are performed as followed.

TCM physicians utilize four methods of diagnosis to examine patients. In congruent with the TCM theory, they dissect the diagnosis into five different-yet-interrelated elemental categories: fire, earth, metal, water, and wood. These five elemental categories refer to specific parts of the human body. Each elemental category can be used to create, counter, and assist other categories. Physicians use this framework as a foundation for their analysis. The final result of the diagnosis is the properties and the severity of a disease. Laboratory testing can confirm and compliment the diagnosis. For treatment, physicians can either prescribe TCM treatment, or refer the patients to western medicine treatment.
The ontology should depict the aforementioned procedure as a part of the knowledge base. The representation of the procedure will increase the ontology’s utility. By enabling the usage of the applications of ontology in the clinical practice settings, the ontology will be better equipped to capture new knowledge.

6. Discussion

The comparisons have revealed many interesting findings. First, at the top level, CACMS’s “Disease” and “Syndrome” terminologies match ICTM’s “Disorder” and “Pattern”. Second, similarities and differences are intermingling within any subclasses. These two findings lead to a conclusion that between CACMS and ICTM ontologies, it is improbable to find any meaningful one-to-one relationship.

Similar to ICD-11, ICTM is a work-in-progress. The ICTM’s TCM terminologies repository will grow larger to accommodate the TCM’s rich knowledge base. ICTM has to account for traditional medicine practices in China, Japan, and Korea. As a result, ICTM employs higher level of abstraction. The abstraction helps to fully capture the traditional medicine practice at a general level. On the other hand, higher level of abstraction will diminish granularity.

We have completed the first stage of our TCM Ontology study. The next stage is to devise specific use cases to identify key point of intersections between these two ontological approaches and suggest a mean to reconcile and/or account for differences. In doing so, particular attention will be paid to the clinical and research implications and uses. For example, the prospect of structured data capture of TCM manifestations, diagnosis, and treatment could provide an unparalleled knowledge and data repository to understand the efficacy of TCM.

We acknowledge that a research study of this magnitude will benefit from various research directions in the future. For instance, we endeavor to employ quantitative methods, conduct empirical observations, and/or use mixed methods. These approaches are important as we delve into more real life validated use cases of the ontology in terms of practical medical practice.

In a related manner, future research will aim to strengthen the connection between the TCM ontology and Health Information Technology and international standards, specifically ICD. This development of TCM ontology could be conducted under the Design Science Research paradigm [9] [10] [17], where research artifacts are created and validated. The artifacts would be the new TCM ontology. This
ontology should undergo testing to confirm its validity. Validation can be done in three directions: one is the utility and consistency to be used in Health Information Technology system, a second is the compatibility of the ontology with international standards, and lastly, the ontology should represent TCM medical practices. The successful creation of this ontology is a step closer to realizing the “Health for One” initiative. For CACMS, it is the breakthrough that allows two outcomes. First, it allows the integration between TCM and international standards, represented by ICTM. Second, it allows for interoperability of ITCM, in conjunction with ICTM, with EHR implementation.

7. Conclusion

Traditional Medicine is one of the most ancient medical and widely used medical traditions and practices around the globe. Yet, advances in Health IT have been focused predominantly on western medicine. The China-US collaboration reported here is a first step to utilize ontology to link TCM with developments in Health IT standards and ontologies. Its advance will not only serve to enhance a strong TCM practice and research platform, but to allow for the integration of TCM and western medicine as practice and research interests dictate and/or allow.

8. References


