Lily Dong

An Ontology of Traditional Chinese Medicine (TCM)

Motivation

Growing up in China, my parents and grandparents used to cure me and my little brother's minor ailments from back pain, allergies, stomachache, and sicknesses like cold and cough using treatments and ingredients from Traditional Chinese Medicine (TCM). I developed an annoying allergy, with a runny nose and sore eyes and throat every day, when I studied high school in Canada. The doctors still could not tell what the allergens were after all the different lab tests. The summer I went back to China, my parents took me to see a doctor who practices TCM. With my mom brewing me a bitter, disgusting taste pot of soup of ingredients that the doctor prescribed every morning, my allergy symptoms gradually went away, magically but as expected. The other treatments that the doctor prescribed were drinking hot water, dressing warm, avoiding spicy and oily food and seafood, and *aijiu* (艾灸) - moxibustion therapy.

Context

TCM is a systematic domain of knowledge and understanding of our body and our relationship with the natural world. It encompasses not only the tangible treatments and ingredients but also principles and theories about the intangible, like the flow of *qi* (气), *yin* and *yang* (阴阳), and the five elements (五行). Each aspect of TCM reflects how people from thousands of years ago understand the body and the natural world. Their wisdom and philosophy are not supernatural beliefs but are deeply rooted in empirical observations and centuries of practical experimentation.

The goal of this project is for me to capture and structure this vast body of knowledge utilizing modern data management tools and techniques to help us better understand this domain. I also hope to use this project to build toward my master's thesis of eventually creating an AI model for providing TCM diagnostic advices for doctors and patients.

Data

The Compendium of Materia Medica -《本草纲目》documents a comprehensive list of medicinal substances derived from plants, animals, and minerals. It contains over 1,800 entries detailing the property, use, and classification of medical ingredients. [Library of Materia Medica](#) is a digitalized database I will use for this project.

Another aspect of data that I want to include in building towards a TCM ontology is *Huangdi Neijing* -《黄帝内经》. It provides a theoretical framework including principles, diagnostic methods, and an explanation of our body system, diseases, and treatments.

Lily Dong

By synthesizing these two sources, the project aims to create a holistic ontology that integrates both the medical components and the theoretical basis of TCM.

Method

How is the class hierarchy determined?

The class hierarchy in this ontology is determined based on the natural organization of Traditional Chinese Medicine (TCM) concepts and their relationships as described in classical texts like *Bencao Gangmu* (Compendium of Materia Medica) and *Huangdi Neijing* (The Yellow Emperor's Inner Canon). I categorized entities of the ontology into three main domains: HumanBodyRelatedEntities (人体), NatureRelatedEntities (自然), and TheoreticalEntities (理论). These categories reflect the holistic framework of TCM: the human body, its connection to the natural world, and the theoretical principles that reflect the wisdom and judgement of how we understand our body and the natural world through TCM. Subcategories within these domains, such as Organs (脏器) under HumanBodyRelatedEntities or Pathogens (致病因素) under NatureRelatedEntities, were derived from TCM's foundational understanding of bodily functions, disease causes, and medicinal resources.

How are the properties determined?

Properties were determined based on the relationships explicitly or implicitly described in TCM texts. These relationships include:

- The relationship between natural pathogens and elements with our body systems (e.g., affects).
- The relationship of how TCM theories relate or could be applied to which part of our body (e.g., appliesTo).
- The relationship of how TCM theories explains our natural world (e.g. explains)
- The relationship between theoretical concepts (e.g., hasReference) or syndromes and symptoms (e.g., hasSymptom).

For example, the property hasIngredient links a formula to its specific components, while manifestsIn connects a syndrome to the body part where it is observed. These properties reflect both practical applications in TCM (e.g., what treatment targets which syndrome) and philosophical connections (e.g., the interplay of Yin and Yang).

Result

The final ontology included a total of 109 classes and 16 object properties. Some classes are asserted, such as the named body parts, medical ingredients, and diseases, etc. Some other classes, like DiagnosticMethods, a "MethodsOfPractice'" that "diagnosesSyndrome some Syndromes" and "identifiesSymptom some Symptoms," are inferred classes based on their definition and role in TCM.

Lily Dong

What were the challenges?

Defining Class vs. Individual:

One of the biggest challenges was deciding whether certain entities, like Heart or Liver, should be modeled as classes (abstract concepts) or individuals (specific instances). As TCM continue to evolve, more concepts and relationships emerge and new knowledge is generated. Listing concepts and entities are a class in the ontology allow more flexibility for future additions of subclasses or new individuals under that class. Because individual is the smallest unit in the ontology, it can be difficult to modify in the future if we want to add children elements to an individual – we would have to change that individual to a class. For that reason, I treated some entities, like heart or liver as classes, rather than as individuals, so that we could add additional unique instances and occurrences of that class.

Deciding the hierarchy of classes and properties:

TCM concepts are interconnected and multilayered, which made it extremely difficult to capturing the richness of the domain while keeping the ontology manageable and organized in a hierarchical structure. To address this problem, I first categorized the ontology into three overarching domains: human, nature, and TCM related knowledge. Then, I referenced TCM textbook, "Basic Theories of TCM" to figure out the subcategories, for example, how the hierarchical structure of body system is defined in TCM. I also referenced Library of Materia Medica for classification of medical ingredients.

Sharing this data with the public

I chose to create a GitHub repository to share this ontology with the public, because a repository allows collaboration, version control, and documentation. I chose the CC-BY-NC 4.0 license (Creative Commons Attribution-NonCommercial 4.0 International) to ensure:

- Attribution: Users must credit work when sharing or adapting it.
- NonCommercial Use: Prevents commercial use of the ontology without permission.

The license was added as metadata in the ontology file and documented in the GitHub repository for transparency.

Lily Dong

References

Li, D. (2001). Zhong yi ji chu li lun (Di 1 ban.). Ren min wei sheng chu ban she.
Li, S., & 李时珍. (1977). Ben cao gang mu (Jiao dian ben.). Ren min wei sheng chu ban she.

Long, H., Zhu, Y., Jia, L., Gao, B., Liu, J., Liu, L., & Herre, H. (2019). An ontological framework for the formalization, organization and usage of TCM-Knowledge. BMC Medical Informatics and Decision Making, 19(S2). https://doi.org/10.1186/s12911-019-0760-9

Mao, Y., & Yin, A. (2009). Ontology Modeling and Development for Traditional Chinese Medicine. 1–5. https://doi.org/10.1109/bmei.2009.5301720

Wu, Z., 吴子腾, Wu, Q., Xu, Z., & Zhang, J. (2014). Zhong yi si da jing dian ming zhu fang ji ji zhu: "Huangdi nei jing" "Shang han lun" "Jin gui yao lüe" "Wen bing tiao bian" (Di 1 ban.). Zhong yi gu ji chu ban she.

本草数典中药知识库. (2025). Sciencereading.cn. https://lmm.sciencereading.cn/#/