

中医药领域本体研究现状*

徐静雯

朱彦

(长春中医药大学医药信息学院 长春 130117) (中国中医科学院中医药信息研究所 北京 100700)

〔摘要〕 对近 10 年来中医药领域本体研究进行文献调研、筛选和统计,梳理中医药领域现有本体相关研究进展、本体主题分布情况,总结当下中医药本体研究存在的不足,并提出建议,以期为我国中医药领域本体构建与应用提供借鉴。

〔关键词〕 中医药; 本体; 研究现状

〔中图分类号〕 R-058 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2023.01.007

Study Status of Ontology in the Field of Traditional Chinese Medicine XU Jingwen, School of Medical Information, Changchun University of Chinese Medicine, Changchun 130117, China; ZHU Yan, Institute of Information on Traditional Chinese Medicine, China Academy of Chinese Medical Sciences, Beijing 100700, China

〔Abstract〕 The paper carries out literature research, screening and statistics on ontology research in the field of traditional Chinese medicine in recent 10 years, sorts out the research progress of existing ontologies and the distribution of ontology topics, summarizes the shortcomings of current research on traditional Chinese medicine ontology, and puts forward suggestions in order to provide references for the construction and application of traditional Chinese medicine ontology.

〔Keywords〕 traditional Chinese medicine (TCM); ontology; research status

1 引言

在生物学大数据研究领域,本体作为支持异构信息标准化和智能分析处理的一种有效手段,已经普遍应用于数据分析、检索、整合、共享与再利用等场景,成为生物学领域资源组织的重要方式^[1]。

与此同时,围绕中医药领域本体展开的研究也在逐渐增多。有学者对其研究成果进行了总结,如李兵等^[2]从中医药领域本体构建的 3 个方面进行综述;田甜等^[3]对本体在中医药领域应用的现状进行梳理等。上述研究用举例说明的形式对 2006—2011 年中医药领域本体研究进行总结。近 10 年来,中医药领域本体的研究与实践取得新的进展,现阶段

〔修回日期〕 2022-06-10

〔作者简介〕 徐静雯,硕士研究生,发表论文 4 篇;通信作者:朱彦,副研究员,硕士生导师。

〔基金项目〕 国家自然科学基金“广义中医经典名方智能辅助遴选系统关键技术研究”(项目编号:82174534);中国中医科学院科技创新工程项目“顶层本体指导下的证候本体构建研究及应用”(项目编号:CI2021A05306);中央级公益性科研院所基本科研业务费专项资金资助“中医药语义信息标准化探索研究”(项目编号:ZZ13-YQ-126);中国中医科学院基本科研业务费自主选题项目“基于开放数据的中成药不良反应知识发现”(项目编号:ZZ150314)。

缺少对中医药领域本体研究现状的系统梳理，以及对本体文献研究类型与本体主题的量化分析，因此亟需全面梳理研究现状，以应对当下大数据环境下中医药知识表示与组织的需求。

在中医药领域本体构建与应用中，有 4 个重要的环节。其一，形式化表达是促进计算机理解与推理的重要前提^[4]。其二，复用顶层本体能够有效实现知识共享与重用^[5]，减少“闭门造车、各自为战”的倾向^[6]，促进本体间的互通。其三，映射或复用周围其他本体可以避免知识的重复性，减少大量本体的重复建设与储存，提升本体间高质量的互操作^[7]。其四，由于提倡可发现（findability）、可访问（accessibility）、可互操作（interoperability）和可重用（reusability）的 FAIR 原则^[8]对科学数据领域的巨大影响，科学数据的可发现性与可访问性已大大提高^[7]，科学数据是否共享逐渐成为中医药领域本体构建与应用的重要内容，在科学数据元中心开放元数据访问的情况下，能够有效促进本体间的关联与融合。

因此，本文首先对 2010—2021 年国内外相关文献进行调研，梳理近 10 年中医药领域本体的研究与实践，对上述 4 个指标情况及本体所涉及的主题分布进行量化统计。其次，对现有中医药领域本体进行归纳总结。最后，根据研究现状提出建议，展望中医药领域本体的未来发展，以期为我国中医药领域本体构建与应用提供借鉴。

2 资料与方法

2.1 文献检索和筛选

首先对中医药本体研究进行文献调研，在中国知网、万方医学网、PubMed、SinoMed 中使用高级检索，检索 2010—2021 年的文献。中国知网检索

式：（（（全文 = 中药）并且（全文 = 本体））或者（（全文 = 中医）并且（全文 = 本体）））（精确匹配）。万方医学网检索式：（（“中医”）AND（“本体”））OR（（“中药”）AND（“本体”））。PubMed 检索式：（（traditional Chinese medicine）AND（ontology））OR（（traditional Chinese medicine）AND（ontological））OR（（traditional Chinese medicine）AND（ontologie））OR（（traditional Chinese medicine）AND（ontological argument））OR（（traditional Chinese medicine）AND（ontologia））。SinoMed 检索式：“中医”[常用字段：智能]AND“本体”[常用字段：智能]OR“中药”[常用字段：智能]AND“本体”[常用字段：智能]。对以上检索到的文献进行人工筛选。

2.2 结果统计

2.2.1 内容分布 依照以上文献检索步骤，共获得文献 126 篇。按照文献研究主题与研究内容进行分类得到本体综述、本体方法学研究和本体构建及应用 3 类。其中综述共计 29 篇，方法学 6 篇，本体构建或应用共计 91 篇，见图 1。

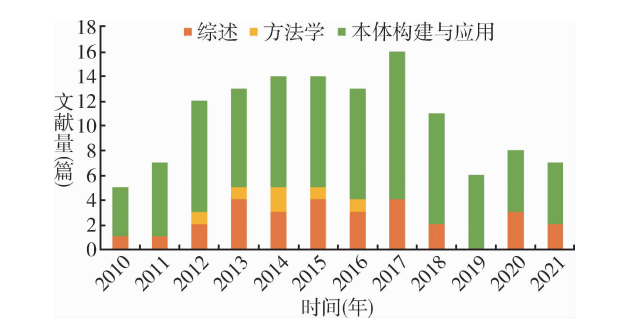


图 1 中医药领域本体文献研究类型年代分布

2.2.2 年代和主题分布 对 91 篇本体构建与应用文章逐一阅读，按照研究内容和主题进行分类，其主要分布主题及发表年代，见表 1。

表 1 本体构建与应用文献按年代和主题统计（篇）

时间（年）	疾病	证候	症状体征	中药	方剂	针灸	中医基础理论	临床诊疗应用型	其他
2010	0	0	0	0	0	0	1	2	1
2011	0	0	0	0	2	1	1	2	0
2012	0	0	2	0	3	0	0	4	0

续表 1

2013	0	2	0	0	1	0	1	3	1
2014	0	1	0	1	0	2	1	3	1
2015	0	1	0	3	0	0	0	3	2
2016	0	0	1	0	2	0	0	4	2
2017	0	0	3	0	0	0	1	6	2
2018	0	0	0	2	0	1	1	4	1
2019	0	0	1	0	1	1	0	2	1
2020	1	0	0	2	0	0	0	2	0
2021	0	0	0	2	0	0	0	3	0
合计	1	4	7	10	9	5	6	38	11

2.2.3 指标分布 根据 91 篇本体构建与应用文献中所构建的本体是否具有形式化表达、是否复用顶层本体、是否映射或复用其他本体、是否支持科学数据共享 4 个指标进行统计,通过对其明确提及或可用判断的指标统计可知,其中具有形式化表达为 58 篇,复用顶层本体为 6 篇,映射或复用领域其他本体为 52 篇,科学数据共享为 4 篇。本文 4 个指标的判定标准如下。(1)形式化表达。文献中所构建的本体是否对相关概念进行定义,即是否使用本体语言进行属性定义。(2)复用顶层本体。文献中所构建的本体是否复用现有生物医学顶层本体。(3)映射或复用其他本体。所构建的本体是否与相关术语集、标准、指南进行映射或复用领域内相关本体。(4)科学数据是否共享。是否有渠道进行下载或浏览本体。

3 分析与思考

3.1 现状统计

基于上述文献调研结果发现,中医药本体研究作为一个细分的研究领域,近年来得到了学界的持续关注。其中,2012—2018 年是一个较长的热点时期,所发表的本体构建与应用的文章每年 10 篇左右;尤其 2012—2016 年,不仅本体构建与应用的研究较多,还进行了一些方法学的研究。2019 年以来则相对平稳,本体构建与应用的文章每年基本在 4 篇左右。

本体按作用可以分为“作为增强版受控术语

集”(enhanced controlled terminology)的参考本体、知识库本体和元数据本体^[9]。中医药领域的本体研究主要集中在前两类,元数据本体也有少量研究^[10-11]。参考本体是传统术语系统的继承和发展,基于文献统计数据,在“理-法-方-药”各个子领域都开展了相关探索和研究。其中中药、方剂子领域,由于其语义模型相对简单,研究相对较多,落地应用也更成熟。相比证候,中医疾病由于缺少实际应用场景,仅见 1 篇基于分类统计的传统医学疾病本体构建研究^[12]。中医基础理论方面,主要有治则治法^[13]和病因病机^[14]的本体研究,未见阴阳、五行等顶层概念的本体建模研究。知识库本体主要集中在临床诊疗应用研究方面,研究者参照中医临床诊疗指南或教科书等共识知识构建,针对专科专病进行探索^[15-16],如中医哮喘领域本体^[17]、银屑病诊疗本体^[18]等,来支持临床辅助决策等场景。其他研究则主要是基于本体在知识分类和映射,尤其是古今中医药知识在连接方面的优势,研究者进行了多个子领域的探索,包括中医骨伤古籍本体^[19],基于本体的中医古籍文献数据库构建^[20],基于针灸本体的古籍文献检索^[21]等。

3.2 分析与思考

3.2.1 大多数中医药领域本体未使用顶层本体 顶层本体可以为中医药领域本体提供其所需的先辈术语,进而提供公共总体框架。中医药领域顶层本体不仅是将领域本体与通用顶层本体相联系的基础,还是中医药领域本体共享性、互操作性的基

础^[22]。通过文献调研,在近 10 年中医药本体构建与应用文献中,仅有 6.59% (6 篇)对顶层本体进行复用,在对顶层本体的复用选择上,生物医学顶层本体方面主要以通用形式化本体 (general formal ontology, GFO)、基本形式本体 (basic formal ontology, BFO)、推荐的顶层融合本体 (suggested upper merged ontology, SUMO) 等顶层本体作为知识表达模型,中医药顶层本体主要以中医药学语言系统 (traditional Chinese medical language system, TCMLS) 为参照。高成勉等^[6]提出要在中医顶层本体的构建基础上加强应用探索,并基于此提出中医药本体构建应以知识的开放共享、互通互联作为本体构建与应用的基本宗旨,连接现有国内外中西医顶层本体,从而实现相互映射、知识整合的目标;董燕^[23]通过对中医临床术语系统、医学系统命名法—临床术语 (systematized nomenclature of medicine—clinical terms, SNOMED CT)、SUMO、BFO 的顶层分类进行比较研究,并尝试将其顶层分类进行对应,分析其顶层本体之间的映射情况。但由于传统的中医学理论有独特的体系架构,无论是哲学起源,还是中医药术语的语义和语境,都与西医的知识体系有着极大差异,因此大多数中医药领域本体在复用生物医学顶层本体时,存在着语义表达的困难和中西概念体系的矛盾^[5]。并且中医药领域本体的构建与中医理论研究关系密切,因此在与生物医学顶层本体对接时存在部分不适用的情况。这些都影响着中医药领域本体对顶层本体的复用。

3.2.2 与已有标准、术语集、本体的协同映射不足 本体作为支持异构数据标准化的重要工具,其可以通过与其他本体或标准之间的协同映射来实现消除数据异质性、整合多来源数据的目的。通过文献调研可知,57.14% (52 篇)本体构建与应用研究采用了映射的方式,但其中部分是对相关标准、指南、教科书等进行映射,而与本体进行直接映射的研究较少。崔家鹏等^[24]在构建脾脏象理论体系知识本体时,为保证术语数据的完整性和准确性,在处理中如发现不同的辞书包含同一名词,则将“释义”合并至同一“全名”条下,但保留各自的出处。崔一迪^[18]在寻常型银屑病中医诊疗本体的构建

中,并未发现有高质量的相关领域本体可复用,在本体映射中主要参照国际疾病分类 (international classification of diseases, ICD),并映射 ICD 11 的疾病编码。这一现象产生的原因与现有中医药领域本体的数据共享不足有关,获取相关领域的本体数据较难,并且本体数据共享之间缺乏有效机制。

3.2.3 中医药领域本体形式化表达有待加强 在本体构建过程中,网络本体语言 (web ontology language, OWL) 作为语义网中的一项核心技术语言,是不可缺少的重要形式化语言。OWL 的一个关键特点是具有描述逻辑 (description logics, DLs)。描述逻辑可以通过计算来实现本体的逻辑推理等功能,其主要是基于一阶逻辑的模型论来进行形式化语义表达。63.74% (58 篇)本体研究已经基于本体编辑软件 Protégé 对 Sub Class 上下级关系和某些属性进行了必要条件定义,但使用 Equivalent To 定义全等关系的本体并不多,在对概念与属性关系的形式化表达上仍有欠缺。孙静^[25]在构建中医症状本体的过程中采用了本体构建 7 步法,但其所构建的顶层概念缺乏清楚的表达边界,如对“症状”的概念,并未清楚表述其内涵与外延。仅通过 Protégé 的注解界面中注明了相关定义,而非形式化表达。并且对于症状类相关属性等并未给出明确概念,因此这类本体难以进行本体推理,仅能作为简单的术语集表示。李明等^[26-27]构建的中医证候本体,综合参照了国家标准与 SNOMED CT 等分类体系与原则,并基于此对中医证候进行分类与映射,但其对证候的概念与定义主要是以特征表述的方式,同时对病因病机等相关概念并未进行形式化定义。由此可见,在本体技术使用方面需要进一步提升,中医药领域本体中形式化定义缺乏,无法有效支持计算机理解概念内涵及属性关系,无法发挥出中医药领域本体的优越性、数据共享以及互操作性,从而导致本体应用困难,难以进一步实现关系推理与智能检索分析等用途。

3.2.4 本体间的科学数据共享仍有待提高 构建中医药本体的主要目的是为了实现在领域内知识的共享和重用,而实现这一目的的前提是本体资源的开放共享。目前本体的存储格式一般为 OWL 或 OBO

两种开放格式,具有数据格式标准化方面的天然优势,能有效支持实现数据的开放和共享。然而,笔者对上述 91 篇本体构建与应用研究进行统计,只有 2.20% (2 篇)^[28-29] 本体构建研究明确提供访问链接及提供 OWL 格式的本体文件下载,2.20% (2 篇) 本体应用研究提供统一资源定位地址 (uniform resource locator, URL) 实现应用系统浏览访问^[30-31]。本体构建后的研究结果无法有效访问,直接阻碍和影响了同行评议的进行以及后续本体的复用。后续对相关研究结果的调研和思考,只能参考文章的描述或截图来进行。提升本体资源的开放与共享,促进本体间的数据交换与融合,是中医药领域本体亟待解决的重要问题之一。

4 建议

4.1 基础理论研究仍有待加深

加深基础理论研究,尤其需要对现有 BFO 等顶层本体的哲学基础进行学习,并开展与中医理论适配的调研,探索一套能融合中医与现代生物医学的顶层本体框架。一方面为中医药领域本体的构建提供顶层理论指导框架;另一方面,也能真正实现与现代生物医学知识的语义互操作。

4.2 与现代生物医学本体社区之间的联系有待加强

加强与现代生物医学本体社区之间的联系,学习借鉴其先进的本体理论、方法和技术。如以开放生物医学本体工场 (the open biological and biomedical ontologies, OBO Foundry) 为代表的生物医学本体开发国际社区,所提出的本体开发原则和流程,具有很高的参考价值,可以快速、直接地为中医药领域本体的构建提供理论和技术方面的参考,有效解决形式化表达不足等技术问题。

4.3 中医药本体顶层框架与合作机制有待开发

广泛开展合作,通过学会等载体,建立中医药领域的本体研究社区,共同研究开发中医药本体顶层框架和合作机制,建立健全数据共享机制,鼓励术语标准以结构化形式进行发布,便于复用和映

射;推动构建在顶层本体指导下开放共享的中医药领域本体,实现不同术语集和标准的融合与协调,从而有效避免重复建设。

5 结语

本文通过对近 10 年中医药领域本体研究现状的梳理、思考及建议,以期为中医药领域本体的研究和应用提供借鉴和参考。随着中医药的现代化呈现出多源异构数据、多学科、多层次、多环节的整合研究趋势,中医药知识的特殊性、复杂性,使其与现代生物医学在哲学层面存在着天然的异构。整合药理学、本草基因学等新兴学科的提出与发展,为收集、组织、集成和共享复杂异构数据提出了新的挑战,也为中医药领域本体研究和应用带来了新的机遇。总之,未来中医药领域本体研究工作任重而道远,后续笔者团队将与相关研究者一起,从激励机制、社区建设、基础理论研究、交流合作与培训等方面,逐步落实这些建议,为中医药信息和智能化做出更多贡献。

参考文献

- 1 赵洁,司莉.国内外生物医学领域本体研究与实践进展[J].数字图书馆论坛,2020(8):7-14.
- 2 李兵,裴俭,张华敏.中医药领域本体研究概述[J].中国中医药信息杂志,2010,17(3):100-101,106.
- 3 田甜,高思华.本体在中医药学领域的应用现状及展望[J].中华中医药杂志,2013,28(6):1832-1834.
- 4 于彤,崔蒙,李敬华,等.中医药本体工程研究现状[J].中国中医药信息杂志,2013,20(7):110-112.
- 5 龙海,贾李蓉,朱玲,等.顶层本体 GFO 在中医药领域本体构建中的应用展望[J].中国中医药图书情报杂志,2016,40(1):15-20.
- 6 高成勉,包含飞,周强.本体构建原则及其在中医顶层本体构建中的应用[J].医学信息,2008(5):581-583.
- 7 卢逸航,李国庆,陈祖刚.科学数据中心间互操作模式研究[J].数据与计算发展前沿,2022,4(1):69-83.
- 8 宋佳,温亮明,李洋.科学数据共享 FAIR 原则:背景、内容及实践[J].情报资料工作,2021,42(1):57-68.
- 9 何勇群,余红,杨啸林,等.本体:生物医学大数据与

- 精准医学研究的基础 [J]. 生物信息学, 2018, 16 (1): 7-14.
- 10 陶金火. 基于语义的中医药数据采集工程及应用平台 [D]. 杭州: 浙江大学, 2011.
- 11 丁侃. 基于知识元信息技术的中医古籍元数据研究 [D]. 北京: 中国中医科学院, 2009.
- 12 任慧玲, 李晓瑛, 王哲, 等. 面向分类统计的传统医学疾病本体构建研究 [J]. 中华医学图书情报杂志, 2020, 29 (11): 1-7.
- 13 侯玉, 张昌林, 车立娟, 等. 构建中医治则治法本体的研究 [J]. 数理医药学杂志, 2010, 23 (5): 603-606.
- 14 汪森, 鲍颖洁, 仇莲胤, 等. 构建中医病因病机本体的研究思路探讨 [J]. 医学信息, 2014 (9): 2-3.
- 15 叶青, 刘丹红, 杨喆, 等. 临床指南模型构建中医学知识的规范化表达方法 [J]. 中国数字医学, 2011, 6 (8): 14-17.
- 16 李新龙, 刘岩, 周莉, 等. 基于指南类文献的辨证论治知识体系顶层本体的构建 [J]. 中医杂志, 2018, 59 (13): 1154-1159.
- 17 朱玲, 李敬华, 于琦, 等. 中医哮喘领域本体的构建 [J]. 中国实验方剂学杂志, 2017, 23 (15): 222-226.
- 18 崔一迪. 寻常型银屑病中医诊疗本体知识库的构建与应用 [D]. 北京: 中国中医科学院, 2020.
- 19 林尔正. 基于中医骨伤古籍本体的语义检索研究 [D]. 福州: 福建中医学院, 2008.
- 20 李明, 周强, 罗晓兰, 等. 中医古籍文献数据库的语义检索系统构建研究 [J]. 时珍国医国药, 2017, 28 (10): 2557-2559.
- 21 纪军, 刘婕, 张馥晴. 基于针灸学本体的古医籍概念检索初探 [J]. 医学信息学杂志, 2011, 32 (11): 48-51, 69.
- 22 龙海, 朱彦. 论 GFO 的基本框架及顶层本体比较研究 [J]. 中国中医药图书情报杂志, 2015 (5): 18-22.
- 23 董燕. 基于本体的中医临床术语体系构建研究 [D]. 北京: 中国中医科学院, 2016.
- 24 崔家鹏, 王彩霞, 袁东超, 等. 基于本体的脾脏象理论知识体系构建研究 [J]. 中华中医药学刊, 2018, 36 (2): 388-391.
- 25 孙静. 中医症状本体构建及其应用示范研究 [D]. 武汉: 湖北中医药大学, 2016.
- 26 李明, 张昌林, 包含飞, 等. 中医证候本体表达模式的构建与应用研究 [J]. 上海中医药杂志, 2013, 47 (1): 7-11.
- 27 李明, 朱邦贤, 周强. 基于 protégé 的中医证候本体构建方法研究 [J]. 数理医药学杂志, 2015, 28 (6): 807-809.
- 28 乔幸潮. 中药材本体构建研究 [D]. 北京: 中国中医科学院, 2020.
- 29 ZHU Y, LIU L, GAO B, et al. TCDO: a community-based ontology for integrative representation and analysis of traditional chinese drugs and their properties [EB/OL]. [2021-11-20]. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8483929/>.
- 30 张小刚. 基于中医药本体的语义关系发现及验证方法 [D]. 杭州: 浙江大学, 2010.
- 31 于彤, 朱玲, 张竹绿, 等. 中医药文献元数据规范在语义网环境中的应用 [J]. 中国数字医学, 2014, 9 (7): 91-93.

《医学信息学杂志》开通微信公众号

《医学信息学杂志》微信公众号现已开通, 作者可通过该平台查阅稿件状态; 读者可阅览当期最新内容、过刊等; 同时提供国内外最新医学信息研究动态、发展前沿等, 搭建编者、作者、读者之间沟通、交流的平台。可在微信添加中找到公众号, 输入“医学信息学杂志”进行确认, 也可扫描右侧二维码添加, 敬请关注!



《医学信息学杂志》编辑部