

本草领域本体构建及其应用^{*}

佟琳¹ 张磊¹ 曾子玲² 叶俊泽¹ 张宇¹ 张华敏³ 杨洪军⁴

(¹ 中国中医科学院中医药信息研究所 北京 100700 ² 中国中医科学院中药研究所 北京 100700
³ 中国中医科学院中医基础理论研究所 北京 100700 ⁴ 中国中医科学院 北京 100700)

〔摘要〕 目的/意义 应用本体技术对本草类古籍知识进行系统整合与表达。方法/过程 采用内容分析法, 借鉴现有中医药领域本体, 形成严谨的本草知识等级体系, 以 Protégé 为本体构建工具, 结合 Neo4j 图数据库进行知识展示及应用, 并通过本体框架术语覆盖率计算和领域专家咨询进行质量评价。结果/结论 构建包括名称、分类、性能等 20 个概念类及等同、归属、治疗等 15 个概念关系的本草领域本体, 术语覆盖率均值达 90.72%, 可用于支持本草领域知识库构建与知识服务。

〔关键词〕 本草知识; 中医药古籍; 本体; 知识图谱

〔中图分类号〕 R-058 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2025.03.011

Ontology Construction of Materia Medica and Its Application

TONG Lin¹, ZHANG Lei¹, ZENG Ziling¹, YE Junze¹, ZHANG Yu¹, ZHANG Huamin³, YANG Hongjun⁴

¹Institute of Information on Traditional Chinese Medicine, China Academy of Chinese Medical Sciences, Beijing 100700, China; ²Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing 100700, China; ³Institute of Basic Theory for Chinese Medicine, China Academy of Chinese Medical Science, Beijing 100700, China; ⁴China Academy of Chinese Medical Sciences, Beijing 100700, China

〔Abstract〕 **Purpose/Significance** To systematically integrate and express knowledge from ancient materia medica literature using ontology construction technology. **Method/Process** The study utilizes content analysis method, reuses existing ontologies in traditional Chinese medicine (TCM), establishes a rigorous knowledge level system for materia medica. Protégé is used as the ontology construction tool, combined with Neo4j graph database for knowledge presentation and application. The quality of the ontology is evaluated by calculating coverage and domain expert consultation. **Result/Conclusion** The ontology for the materia medica field is constructed, including 20 concept classes such as names, classifications and properties, and 15 relationships like equivalence, belonging, and treatment. The average coverage of the ontology framework terms is 90.72%. It can support the construction of knowledge bases and knowledge services in the materia medica field.

〔Keywords〕 materia medica knowledge; ancient books of traditional Chinese medicine; ontology; knowledge graph

〔修回日期〕 2024-12-13

〔作者简介〕 佟琳, 博士, 副研究员, 发表论文 60 余篇; 通信作者: 张华敏, 杨洪军。

〔基金项目〕 国家重点研发计划项目 (项目编号: 2023YFC3502900); 中央级公益性科研院所基本科研业务费资助项目 (项目编号: ZZ13-YQ-131)。

1 引言

古代本草类医籍是中医药学的宝贵财富，记载着丰富的药物资源和医学智慧^[1]。但其体裁多样，书体结构、字词术语与现代有所不同^[2]，导致传统的本草知识表达方式难以被现代科技有效挖掘和利用，亟须找到有效的知识组织方式系统整理和呈现珍贵的本草知识。本体（ontology）源于哲学领域，旨在揭示事物的本质^[3]。目前本体作为一种语义和知识层次描述信息系统的概念建模工具，已被广泛应用于知识工程、系统建模、信息处理、自然语言理解等领域^[4]。中医药领域学者在“理、法、方、药”各子领域开展探索研究，知识库本体多集中于临床诊疗应用^[5]。本体技术为中医知识框架构建提供有力支持，促进中医知识复用和共享。本研究结合中药学理论与本草古籍特点，利用本体技术对古籍中本草知识体系进行表达，构建一个开放、语义表达良好、可验证的本草领域本体，以期为本草领域知识管理、图谱构建、辅助决策及智能问答系统开发等应用提供基础，助力本草知识的传承和创新。

2 资料与方法

2.1 数据来源

数据来源为 1911 年及以前本草类古籍文献 50 种，涵盖《神农本草经》《经史证类备急本草》《本草纲目》等从汉至清代各时期代表性本草著作，共收集约 200 万字文本资料，经过初步筛选，最终采用语料体量约 10 万字，部分语料经过预处理和标注。采集数据筛选标准：排除本草相关性不明的古籍条文；排除文献中表述本草内容但未涉及基本理论、实践应用实质内容的条文；排除内容庞杂无重点论述本草知识的条文；排除语义不完整或无上下文信息的条文。

2.2 工具

采用 Protégé 为核心工具构建本草领域本体。Protégé 是一款在生物医学领域广泛应用的本体构建

与编辑软件，支持本体复用并能提供推理机制和可视化界面，可提升本体构建的效率和准确性^[6-7]，为本草领域知识表示和推理提供支持。采用 Neo4j 图数据库进行本草知识的实例展示和应用。Neo4j 图数据库能高效处理节点与关系构成的复杂网络结构，通过 Cypher 语言可实现数据的增、删、改、查，为知识的存储、管理和应用提供支持^[8]。

2.3 方法

采用本体构建法建立本草知识图谱模式层，形式化表示概念类与关系，规范数据层实体类型、关系和属性。先采用内容分析法梳理古籍本文中的概念类与关系，自上而下构建本体，再通过数据层归纳知识，自下而上完善。使用 Protégé 5.5.0 构建本草知识图谱框架，选择 OWL 为描述语言，借鉴领域本体 7 步构建流程^[9]，本草知识本体构建步骤，见图 1。

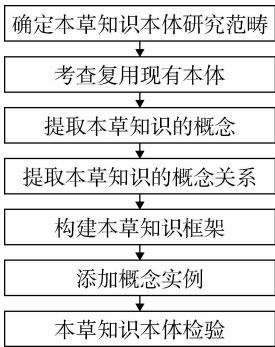


图 1 本草知识本体构建步骤

2.3.1 确定本草知识本体研究范畴 选择 1911 年及以前著述出版的代表性本草古籍专著 50 种，构建以本草知识基础理论、实践应用以及古籍文献为内容的语料库。尽量保证本体的研究范畴能够涵盖和反映本草古籍的概念类与关系，全面系统展示本草知识体系。

2.3.2 考查复用现有本体 国外生物医学领域本体数据库有近百种，如统一医学语言系统、系统化临床医学术语集、基因本体、人类表型本体、疾病本体和药物本体等^[10-11]。当前中医药领域本体相对较少，包括中医药学语言系统语义网络框架、中医临床术语系统^[12-13]等。因上述本体涵盖本研究

部分知识，复用其中相关基本概念和语义关系，促进知识有效重用，便于知识共享。

2.3.3 提取本草知识的概念 概念^[14]是构成本体最基本的知识要素，术语^[15]是特定学科领域概念的称谓集合。遵循客观、合理、准确、完整原则提取本草知识概念，将符合范畴的代表性本草古籍原文整理成文档形式，用于概念提取及管理。由两名专业人员提取，字段包括提取概念及来源，发生分歧时由领域专家判断。整理本草术语时，因古汉语省略需研读原文补全，按原则提取术语，保留古籍原貌与术语逻辑关系，辨析语义，明确上位术语并予以定义。

2.3.4 提取本草知识的概念关系 遵循准确定义、用词简洁、范围明确、合理适用的原则，参考相关领域本体及中医药语言系统关系的研究成果，复用如《健康信息学 中医药学语言系统语义网络框架》中的概念关系。对于新增概念关系的提取，基于本草文献结构与内容特征分析，剖析概念之间语义关系，并结合专家经验，确保符合本草学科特点并满足实际应用需求。

2.3.5 构建本草知识框架 明确本草概念等级体系，构建严谨知识框架。先依据本草知识特性与专家建议，定义本草知识核心分类框架；再基于此框架，对古籍中本草知识进行系统组织分类；通过精细层级划分与逻辑关系梳理，确保本草知识等级体系科学实用，为后续研究与应用奠定基础。

2.3.6 添加概念实例 实例^[16]是本体具体的存在，是类的说明。将语义分析抽取的实例添加至概念类，部分实例如下。基本理论类：名称实例为杜仲、阿胶；分类实例为上品、木部；性能实例为辛、平；地域实例为汉中、上党；采收实例为夏、采皮；鉴定实例为皮中有丝、木高数丈。临证实践类：功用实例为补肾、安胎；主治实例为胎漏、崩中；组方实例为大造丸、补肾汤；炮制实例为炒、

研末；用法实例为白汤下、酒服；用量实例为八分、一两；禁忌实例为肾虚火炽、肝肾阴虚。文献来源类：古籍名称实例为《本草纲目》《肘后备急方》；著者实例为李时珍、朱丹溪；著作年实例为219 年、1593 年；著作朝代实例为宋、元；古籍版本实例为明万历二十一年金陵胡承龙刻本、明万历二年李栻刻本。

2.3.7 本草知识本体检验 为了确保本草本体框架的准确性与实用性，计算本体框架覆盖率^[17]。一是过滤掉本研究构建本体所用的古文献内容。二是将剩余文献 2 000 条编号。三是根据编号随机选取文献，每次选 20 条，共选 3 次。四是由专业背景人员从中提取术语，与本草知识本体框架分类进行人工比对，判断本体是否覆盖这些术语。五是统计 3 次抽取的术语总数、术语覆盖数和覆盖率。采用德尔菲法^[18]邀请领域专家评价本草知识本体框架的可行性。一是编制问卷，包括专家基本情况、背景介绍与填写说明、专家对语义类型与语义关系重要性的 5 级评价、专家判断能力 4 部分。二是制定专家选取原则，包括：专家对研究主题具有权威性，分布均衡具有代表性；主要领域为文献学、中药学、方法学；高级职称，愿意参与，熟悉德尔菲法。三是指标筛选原则：重要性评分 > 4.0；变异系数 CV < 0.25；结合专家意见。四是结果统计，使用 Access 2016 构建问卷结果数据库，使用 SPSS 22.0 软件进行分析。

3 结果

3.1 本草本体构建结果

3.1.1 本草知识概念类 通过本草文献调研，解析术语内涵与外延，比较多来源概念定义，确定上位术语概念定义，抽取规范本草知识体系概念类型，包括名称、分类、性能等。概念类及定义，见表 1。

表 1 概念类及定义

序号	概念类型		定义	属性或实例
1	名称	指中药名称		正名、别名、俗名、代称
2	分类	指记载中草药的古籍所采用的分类方法		三品、来源、自然属性、功能分类
3	性能	指药材本身的性能		阴阳、五行、四气、五味、归经、升降、毒性、动静、缓峻

续表 1

序号	概念类型	定义	属性或实例
4	药用部位	指各种动植物药可以作为药材使用的部位	根、根茎、叶、花、果实、种子、皮、骨、甲片
5	地域	指中药的出产地	产地、道地
6	采收	指中草药的采摘收集	采收时间、采收方式
7	鉴定	指中草药鉴别相关方法与内容	色泽、气味、形状、质地、辨伪、质量、类药
8	功用	指药物产生的功能和用处	功效、副作用
9	主治	指药物的主要疗效	主证、主病
10	组方	指中医治疗疾病时使用的处方名称	方剂
11	炮制	指加工中草药的方法	制法、器具、炮制时间、辅料、贮藏、禁忌
12	用法	指中草药的使用方法	内服、外用
13	用量	指中草药在一定时间使用的数量	两、钱、分
14	禁忌	指用药方面要注意用药禁忌及药物之间的反应	配伍禁忌、饮食禁忌、人群禁忌
15	古籍分类	指中医古代书籍的分类，参照《中医药古籍分类标准》	子部医家、本草类、历代本草、食疗本草
16	古籍名称	指古籍刊刻的名称	《本草纲目》
17	著者	指著书或作文的人	李时珍
18	著作年	指书籍形成的时间	1593
19	著作朝代	指书籍形成的朝代	明
20	古籍版本	指古籍在流通过程中形成的不同版本	明万历二十一年金陵胡承龙刻本

3.1.2 本草知识概念关系（表 2）

表 2 概念关系及定义

序号	主体	关系	客体
1	本草	等同	名称
2	本草	归属	分类、古籍分类
3	本草	属性	性能
4	本草	配伍	本草
5	本草	空间上相关	地域
6	本草	产生	功用
7	本草	治疗	主治
8	本草	组成	组方
9	本草	作用	归经、药用部位
10	本草	方法	炮制、采收、鉴定、用法
11	本草	剂量	用量
12	本草	禁忌	禁忌
13	本草	时间上相关	炮制、采收、用法、著作年、著作朝代
14	本草	记载	古籍名称、古籍版本
15	本草	论述	著者

一方面，参考借鉴相关领域本体及中医药语言系统等，复用现有中药概念关系；另一方面，分析本草文献结构与内容特征，遵循准确定义、用词简洁、范围明确、合理适用的概念关系提取原则，剖析语义关系，并结合专家建议设定新增概念关系^[19]。最终确定以本草为主体的概念关系并描述定义。

3.1.3 本草知识框架 构建本草概念的等级体系，结合本草知识分类框架所定义的概念，对古籍中本草知识进行组织分类，模式层包括上述 20 个类与 15 个关系，支持本草知识组织表达与发现。（1）基本理论类。阐释本草相关基本概念、基本知识和理论等。包括：名称、分类、性能、配伍、地域、采收、鉴定等。（2）临证实践类。反映本草与中医临床治疗实践。包括：功用、主治、组方、炮制、用法、用量、禁忌等。（3）文献来源类。体现本草知识的古代文献出处等。包括：古籍名称、著者、著作年、著作朝代、古籍版本等。

论》知识图谱^[25]，以及将针灸本体应用于古籍文献检索等领域^[26]，为后续研究提供方法和技术借鉴。本研究首次尝试构建本草领域本体，对 1911 年及以前的本草类古籍文献知识进行系统整合与表达，通过有序整合分散在众多古籍中的本草知识，让复杂晦涩的知识变得条理清晰，便于检索和利用，为传统中药知识有效组织与挖掘应用提供支撑。

4.2 本草领域本体构建的方法与挑战

本研究采用内容分析法，复用借鉴现有中医药领域本体，构建本草知识框架，添加概念实例，形成严谨的本草知识等级体系。以 Protégé 为本体构建工具，通过计算本体覆盖率和专家咨询保障本体质量，结合 Neo4j 图数据库展示应用。构建的本草领域本体含 20 个概念类和 15 个概念关系，经检验本体框架术语覆盖率达 90.72%，表明其具有较高的准确性和完整性。基于此开发中药杜仲知识图谱原型，实现本草知识可视化，为中药知识库和知识服务提供支持。构建过程中存在挑战，如本草知识概念关系复杂，层级与交叉关联并存，需深入理解才能准确提取，保障语义表达和推理能力；古籍字词与现代不同，增加概念提取难度，为此采用标注、语义分析等方法，参考现代标准、词典及专家共识，确保提取准确。未来，本草领域本体构建关键在于融合古今知识，提升完整性和应用价值。

4.3 本草领域本体构建的意义与应用前景

本草领域本体构建及其应用研究在中医药学领域具有双重价值。一是系统整合本草学知识体系，助力挖掘其科学与实用价值，推动中药学传承发展。通过本体构建，按性味、功效、形态等维度清晰呈现本草古籍中药，有助于理解古代中医药理论。如本研究发现不同古籍药物分类存在差异，体现其多样性与复杂性，侧面证实中药理论在历史进程中的演变，从知识组织层面为理论研究提供支撑。二是从古籍知识中挖掘药物功效、主治病症等信息，为现代药物研究提供资源，为新药研发提供

思路。如本研究提取古籍中药物性味与功效关系，显示二者关联显著，这种关联有助于解释药物作用机制，可为药物研发提供新视角。本草知识本体构建有望搭建中药知识传承创新桥梁，促进古今知识融合，在智能诊疗、精准用药等领域发挥积极作用。

5 结语

本研究聚焦本草领域，整合古籍知识构建本草领域本体，开发中药杜仲知识图谱原型，初步实现本草知识可视化，为中药创新研究提供新视角。但本草知识体系庞大且复杂，部分概念需进一步扩展，关系和实例尚不全面，如何更有效地整合和表达这些知识仍需探索。未来将利用人工智能和大数

据技术深化本体构建，完善优化模型，探索其在中药创新研究等方面的应用潜力，推动本草学现代化发展。

作者贡献：佟琳负责技术实施、论文撰写与修订；张磊、曾子玲、叶俊泽、张宇负责数据处理；张华敏、杨洪军负责技术指导、论文修订。

利益声明：所有作者均声明不存在利益冲突。

参考文献

1 马乾, 岳朗, 周艳红, 等. 新中国成立以来本草文献语言研究述评 [J]. 陕西中医药大学学报, 2024, 47 (1): 143 - 149.

2 李飞跃. 书同文字与再造书契——论古籍数字化时代的字符统一与文本规范 [J]. 北京师范大学学报 (社会科学版), 2023 (5): 127 - 141.

3 安小米, 张红卫, 魏玮, 等. 跨领域本体标准连续体概念体系研究 [J]. 情报工程, 2023, 9 (6): 3 - 19.

4 李爱华, 徐以则, 迟钰雪. 本体构建及应用综述 [J]. 情报理论与实践, 2023, 46 (11): 189 - 195.

5 徐静雯, 朱彦. 中医药领域本体研究现状 [J]. 医学信息学杂志, 2023, 44 (1): 36 - 41.

6 Protégé [EB/OL]. [2024 - 09 - 01]. <http://protege.stanford.edu>.

7 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.
- 8 Neo4j [EB/OL]. [2024-09-01]. <http://neo4j.com>.
- 9 黄贺瑄, 王晓燕, 顾正位, 等. 医学知识图谱构建技术及发展现状研究 [J]. 计算机工程与应用, 2023, 59 (13): 33-48.
- 10 张璐璐, 杨晟, 史涪仁, 等. 本体支持的生物医学领域元数据异质性与可兼容性研究 [J]. 中国生物医学工程学报, 2019, 38 (3): 324-331.
- 11 张庆, 吕少妮, 轩扬. 本体在生物医学领域中应用研究热点分析 [J]. 医学信息学杂志, 2019, 40 (1): 63-67.
- 12 于彤, 崔蒙, 李海燕, 等. 中医药学语言系统的语义网络框架: 一个面向中医药领域的规范化顶层本体 [J]. 中国数字医学, 2014, 9 (1): 44-47.
- 13 贾李蓉, 李海燕, 刘静, 等. 中医药学术语系统研究概述 [J]. 中国中医药图书情报杂志, 2015, 39 (5): 7-10.
- 14 李思思, 燕燕, 夏书剑, 等. 中医汤剂知识图谱的构建与样例查询方法研究 [J]. 中华中医药学刊, 2024, 42 (8): 31-36.
- 15 叶其松. 汉语术语学术语体系构建的“不变”与“变”——从国际术语标准 ISO1087 说起 [J]. 中国科技术语, 2024, 26 (1): 54-62.
- 16 唐晓波, 王琼赋, 牟昊. 基于词共现与词向量的概念层次关系自动抽取模型——以学术论文评价领域为例 [J]. 情报科学, 2022, 40 (10): 3-11.
- 17 许冬冬, 林惠, 段会龙, 等. 健康行为改变干预本体的构建与应用 [J]. 中国生物医学工程学报, 2023, 42 (1): 74-81.
- 18 高峰, 刘晴, 靳英辉, 等. 基于知识微调和信息融合的医学指南知识抽取 [J/OL]. 武汉大学学报 (理学版), 1-11 [2024-06-19]. <https://doi.org/10.14188/j.1671-8836.2024.0032>.
- 19 张宇, 张舒琪, 佟琳, 等. 基于中医药学语言系统的中医药优势病种术语集构建研究 [J]. 中国中医药信息杂志, 2024, 31 (5): 36-41.
- 20 祝振媛, 李广建. 多视角下的情报分析模型研究综述 [J]. 图书情报工作, 2019, 63 (19): 136-147.
- 21 范瑾研. 中医小儿哮喘诊疗知识本体构建研究 [D]. 长春: 长春中医药大学, 2024.
- 22 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.
- 23 张萌. 中医古籍疫病方剂知识库构建研究 [D]. 长春: 吉林大学, 2024.
- 24 李贺, 祝琳琳, 刘嘉宇, 等. 基于本体的简帛医药知识组织研究 [J]. 图书情报工作, 2022, 66 (22): 16-27.
- 25 林睿凡. 基于本体方法构建唐本《伤寒论》知识图谱 [D]. 北京: 中国中医科学院, 2022.
- 26 孙华君, 李海燕, 贾李蓉, 等. 基于国际标准 ISO/TS16843-6: 2022 的针刺效应本体构建研究 [J]. 中国针灸, 2023, 43 (1): 73-78.

(上接第 41 页)

- 9 BELLO A, NG S C, LEUNG M F. A BERT framework to sentiment analysis of tweets [J]. Sensors, 2023, 23 (1): 506.
- 10 吴雪华, 毛进, 陈思菁, 等. 突发事件应急行动支撑信息的自动识别与分类研究 [J]. 情报学报, 2021, 40 (8): 817-830.
- 11 胡昊天, 邓三鸿, 王东波, 等. 情报学视角下的预训练语言模型研究进展 [J]. 图书情报工作, 2024, 68 (3): 130-150.
- 12 宫小翠, 安新颖, 单连慧. 基于 Labeled LDA 主题模型的医学文献自动分类法 [J]. 中华医学图书情报杂志, 2018, 27 (10): 53-58.
- 13 YAO D, YANG J, ZHANG X. Feature selection algorithm based on random forest [J]. Journal of Jilin university (engineering and technology edition), 2014, 44 (1): 137-141.
- 14 张梦芸, 丁敬达. 面向短文本分类的语义增强研究 [J]. 图书情报工作, 2023, 67 (9): 4-11.
- 15 MeSH [EB/OL]. [2024-05-20]. <https://meshb.nlm.nih.gov/>.
- 16 唐焕玲, 卫红敏, 王育林, 等. 结合 LDA 与 Word2vec 的文本语义增强方法 [J]. 计算机工程与应用, 2022, 58 (13): 135-145.
- 17 罗宏宇, 刘伟. 基于语义层级细粒度的海量文献标引研究 [J]. 情报理论与实践, 2024, 47 (5): 194-203, 193.
- 18 徐璐, 卢小宾, 杨冠灿. 金融科技专利识别与分类方法构建及应用 [J]. 图书情报工作, 2020, 64 (11): 87-95.