

Protégé 在构建中医药本体中的应用*

杨郑子衿

(1 中南大学湘雅二医院 长沙 410013 2 中南大学湘雅三医院 长沙 410013 3 中南大学生命科学学院 长沙 410013
4 医学信息研究湖南省普通高等学校重点实验室(中南大学) 长沙 410013)

徐 倩

(1 中南大学湘雅三医院 长沙 410013 2 中南大学生命科学学院 长沙 410013
3 医学信息研究湖南省普通高等学校重点实验室(中南大学) 长沙 410013)

王安莉

(1 中南大学湘雅三医院 长沙 410013 2 医学信息研究湖南省普通高等学校重点实验室(中南大学) 长沙 410013
3 移动医疗教育部-中国移动联合实验室 长沙 410013)

罗爱静

(1 中南大学湘雅二医院 长沙 410013 2 中南大学湘雅三医院 长沙 410013 3 中南大学生命科学学院 长沙 410013
4 医学信息研究湖南省普通高等学校重点实验室(中南大学) 长沙 410013
5 移动医疗教育部-中国移动联合实验室 长沙 410013)

〔摘要〕 以分析已有的中医药基本理论和知识来源为基础,运用 Protégé 软件将中医药知识进行系统的整理、表达,定义知识语义、属性关联并扩展实例,为中医用药的应用研究提供知识推理框架,提高中医药领域知识信息共享性和可用性。

〔关键词〕 本体; 中药; Protégé; 知识表示

〔中图分类号〕 R-058 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2021.06.006

Study on the Building of Traditional Chinese Medicine Ontology by Using Protégé YANGZHENG Zijin, 1The Second Xiangya Hospital, Changsha 410013, 2The Third Xiangya Hospital, Changsha 410013, 3School of Life Sciences, Central South University, Changsha 410013,

〔收稿日期〕 2020-11-12

〔作者简介〕 杨郑子衿, 硕士研究生, 发表论文1篇; 通讯作者: 罗爱静, 博士, 教授。

〔基金项目〕 长沙市科学技术局项目“长沙市智能辅助诊疗工程技术研究中心”(项目编号: kq1901133); 移动医疗教育部-中国移动联合实验室项目“实用型医学知识图谱构建及推理关键技术研究”。

4College of Hunan Province, Key Laboratory of Medical Information Research (Central South University), Changsha 410013, China; XU Qian, 1The Third Xiangya Hospital, Changsha 410013, 2School of Life Sciences, Central South University, Changsha 410013, 3College of Hunan Province, Key Laboratory of Medical Information Research (Central South University), Changsha 410013, China; WANG Anli, 1The Third Xiangya Hospital, Changsha 410013, 2College of Hunan Province, Key Laboratory of Medical Information Research (Central South University), Changsha 410013, 3Mobile Health Ministry of Education – China Mobile Joint Laboratory, Changsha 410013, China; LUO Aijing, 1The Second Xiangya Hospital, Changsha 410013, 2The Third Xiangya Hospital, Changsha 410013, 3School of Life Sciences, Central South University, Changsha 410013, 4College of Hunan Province, Key Laboratory of Medical Information Research (Central South University), Changsha 410013, 5Mobile Health Ministry of Education – China Mobile Joint Laboratory, Changsha 410013, China

[Abstract] Based on the analysis of the existing basic theories and knowledge sources of Traditional Chinese Medicine (TCM), the paper systematically sorts out and represents TCM knowledge by using Protégé software, defines knowledge semantics, attribute association and extends instances, provides a framework of knowledge reasoning for the application study of TCM medication and improves the sharing and availability of knowledge information in the field of TCM.

[Keywords] ontology; Traditional Chinese Medicine (TCM); Protégé; knowledge representation

1 引言

近年知识图谱与本体构建、中医药领域知识表示框架研究较为热门,目前最系统化的中医药术语标准化体系有中医药学语言系统(Traditional Chinese Medicine Language System, TCMLS),国家知识基础设施(National Knowledge Infrastructure, NKI)的中医药本体构建^[1],中医药临床术语集(Traditional Chinese Medicine Clinical Terms, TCMCT)等大型术语系统。Zhou X, Wu Z 和 Yin A 等^[2]以已有中医药知识体系为基础,研究开发了支持中医语言知识检索、基于概念的信息检索和信息集成的中医语言系统(Unified Traditional Chinese Medical Language System, UTCMLS)本体应用程序。运用 Protégé 软件构建本体是直观展示中医药知识体系的重要方式之一,本文以分析已有的中医药基本理论和知识来源为基础,通过 Protégé 编辑器将中医药知识进行系统的整理、表达,定义知识语义、属性关联并扩展实例,实现中医药信息知识共建、共享和重用。

2 中医药本体知识相关概念及联系

2.1 中医药本体概念

本体是共享概念模型的明确形式化规范说明,具有强语义特性,能够以直观、图形化的方式来展示领域知识,提高异构系统之间的互操作性,有助

于知识共享^[3]。由于中医药领域元数据存在完整性不足、模糊不清等问题,需要构建中医药物本体,以便系统地归纳整合数据,有利于领域知识共享和重用。本体知识表示过程是构建知识图谱的基础,在中医知识图谱的构建上,已有学者开展相关理论研究和实际探索,目前医学中已有的本体数据有 MeSH、OMIM、Gene Ontology、HPO 等^[4]。

2.2 中医药本体构建知识来源

在本体构建方法上,目前研究多以 TCMLS 为框架,将《中国中医药学主题词表》中的关键主题词作为术语体系和分类依据,针对中医药学知识表达和知识推理进行探索^[3-6]。已有的中医领域本体较多,由于中医领域的复杂性,现有研究侧重点各有不同,主要集中于中医病证^[7]、中医证候学^[8]、治则治法^[9]和方剂学^[10]等方面的知识表达。Miao F、Liu H 和 Huang Y 等^[11]通过整合相关疾病、症状和处方数据资源构建本体,提取核心概念、概念描述、两个实体之间的关系以及中医药领域中关系的限制约束,借助 Protégé 软件从中医药领域本体中获取语义关系。本研究知识来源主要为《中医药学主题词表》、TCMLS 及 2020 年版《中华人民共和国药典》(以下简称《中国药典》)^[12]。为实现描述语言转换,引入《中医药主题词表》,需要在领域本体中切合专业领域研究,因此依据专业知识进行再次重构,使其具备更强的学术性、专业性。TCMLS 主要分为语义网络和基础词库两部分,其中语义网络部分定义 128 种语义类型,其框架构成系统顶层架

构,依据该语义体系,为便于理解,将概念与临床实际相结合,对中药药物语义类型构建进行整理分类,根据《中国药典》对中药系统概念进行具体内涵划分,将其单列为“中成药”“中药材饮片”及“方剂”以增强内涵逻辑性和描述性^[13],见图1。

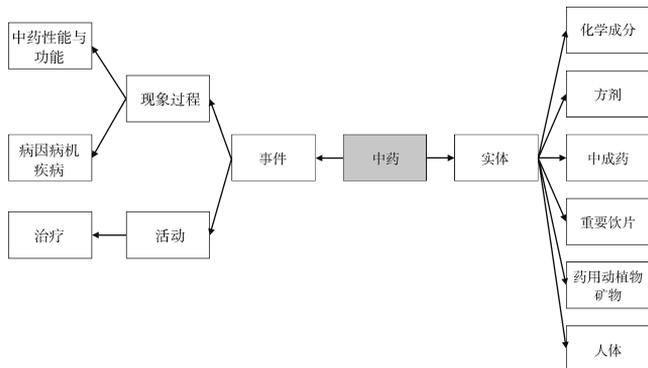


图1 中药语义类型简化顶层本体构建

3 中医药本体构建

3.1 本体构建方法

选取由斯坦福大学提出的“七步法”来构建中医药本体,具体步骤,见图2。首先确定中医药领域本体构建专业内容范畴;对现有中医药本体领域研究进行文献调研,考查是否有可以复用的本体框架;依据《中医药主题词表》列出中医药领域主要术语;将定义的术语映射到类中自上而下构成等级体系;明确定义中医药体系中类的属性;依据中医药领域数据来源定义属性的分面;以方剂为例建立实例。

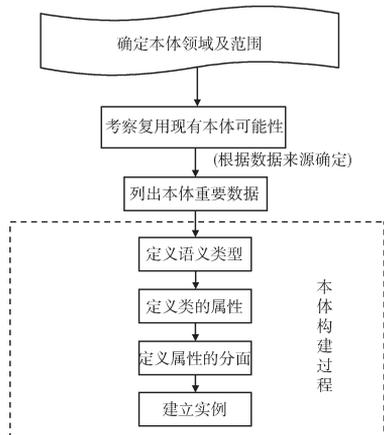


图2 “七步法”流程

3.2 中医药本体类的构建

选择标签 class,创建基本类结构,基于 TCMLS 简化中药本体顶层系统构建及本体构建知识来源,将主体中药主要分成“事件”与“实体”两大类,由抽象概念映射到具体概念中:“事件”下分2级类“活动”与“现象过程”,“实体”下分2级类“中成药”“中药材(饮片)”“人体”“化学成分”“方剂”和“药用动植物矿物”,最终建立的类目体系,见图3。



图3 Protégé 类目体系

3.3 中医药本体属性定义

对构建的类进行对象属性和属性的分面定义,本体中类与类之间的二元关系有两种类型的属性:对象属性 (Object Properties) 和数据型属性 (Data Properties)^[10]。选择自上而下的方式定义层级结构,“上下位父子关系”用“is...a”来构建层次关系,是有指向性的单向关系层级,在其他对象属性中“基源”和“是...的基源”,“组成”和“由...组成”两组互逆,即逆属性 (Inverse Properties) 关系,是目标概念与源概念之间的双向关系,便于后期推理,由于部分类之间的二元关系具有及物 (Transitive) 属性,即上下传递性,因此在关系中设置传递性选项。数据属性分类依据《中医药学主题词表》及《中国药典》,例如以《中国药典》中

的药物“刀豆”为例，其可提取数据属性有：“性状”“鉴别”“炮制”“归经”“用法用量”“功用”“贮藏”，该药材饮片在中医临床术语系统中的逻辑性定义，见图 4。以此方法进行具体层面的属性归纳总结，中医药用药涉及的重要属性，见表 1。在本文的中医药本体构建中，所有术语都有属性定义

的差异，即在一个父词术语下属的子类中应与其他相关的子类术语有区别，因此在考虑类的内部结构时要考虑类之间关系的精确定义。通过确定上述属性定义，完成中医药用药语义关系构建，中药本体构建主干部分，见图 5。中药本体全貌结构局部放大图，见图 6。

刀豆(药材)
+甘(药味)
+温(药性)
+归胃、肾经(归经)
+温中、下气、止呃(功效)

=刀豆饮片

图 4 刀豆饮片逻辑性定义

表 1 中药本体类之间的属性关系

| 主体 | 关系描述 | 客体 | 关系类型 | |
|------|--------|---------|------|------|
| 中药 | 主治 | 病因病机疾病 | 对象属性 | |
| | 具有药性 | 性能 | 对象属性 | |
| | 功效 | 功能功效 | 对象属性 | |
| | 基源 | 药用动植物矿物 | 对象属性 | |
| | 影响 | 人体 | 对象属性 | |
| | 病因病机疾病 | 进程 | 人体 | 对象属性 |
| | | 特征 | | 数据属性 |
| 病机 | | | 数据属性 | |
| 发病类型 | | | 数据属性 | |
| 中成药 | 发病因素 | | 数据属性 | |
| | 是…的成分 | 中药材（饮片） | 对象属性 | |
| | is_ a | 中药 | 对象属性 | |
| | 名称 | | 数据属性 | |
| | 拼音 | | 数据属性 | |
| | 性状 | | 数据属性 | |
| | 鉴别 | | 数据属性 | |
| | 剂型 | | 数据属性 | |
| | 贮藏 | | 数据属性 | |

续表 1

| | | | |
|-------------|-------|----------|------|
| | 用法用量 | | 数据属性 |
| 中药材 (饮片) | 是…的成分 | 化学成分 | 对象属性 |
| | is_ a | 中药 | 对象属性 |
| | 组成 | 中成药 | 对象属性 |
| | 组成 | 方剂 | 对象属性 |
| | 配伍 | 中药材（饮片） | 对象属性 |
| | 中药名 | | 数据属性 |
| | 拉丁名 | | 数据属性 |
| | 拼音 | | 数据属性 |
| | 别名 | | 数据属性 |
| | 形状 | | 数据属性 |
| 方剂 | 用法用量 | | 数据属性 |
| | 贮藏 | | 数据属性 |
| | 鉴别 | | 数据属性 |
| | 质地 | | 数据属性 |
| | 颜色 | | 数据属性 |
| | 浸出物 | | 数据属性 |
| | 制剂 | | 数据属性 |
| | 炮制 | | 数据属性 |
| | 由…组成 | 中药材（饮片） | 对象属性 |
| | 主治 | 疾病、症状、证候 | 对象属性 |
| 药用动物 | 由…组成 | 中药 | 对象属性 |
| | 方剂名 | | 数据属性 |
| | 用法 | | 数据属性 |
| | 功用 | | 数据属性 |
| | 方解 | | 数据属性 |
| | 证治机理 | | 数据属性 |
| | 配伍特点 | | 数据属性 |
| | 是…的基源 | 中药 | 对象属性 |
| | 名称 | | 数据属性 |
| | 植物矿物 | 科属 | 数据属性 |

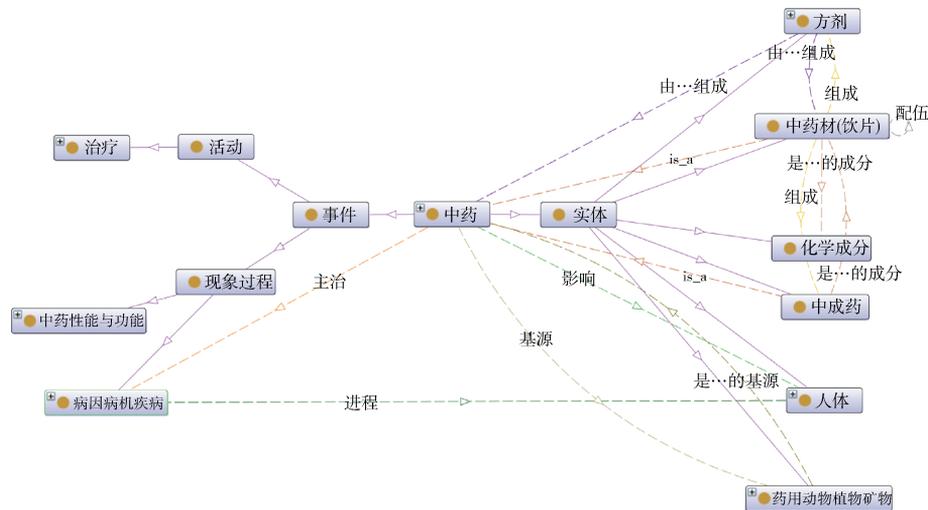


图5 中药本体主体结构

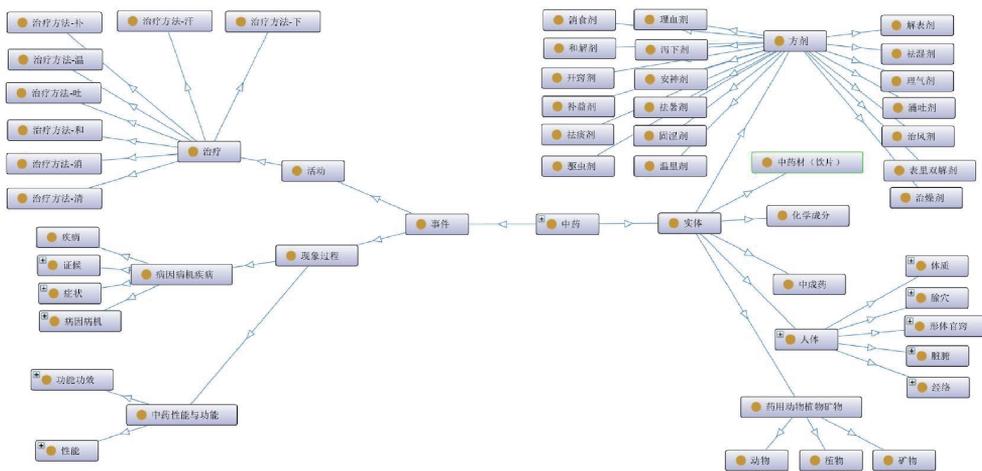


图6 中药本体全貌结构 (局部放大)

3.4 创建实例

3.4.1 实例1 在 Protégé 中, 为解释药物与病因病机之间的关系, 将方剂与症状属性相关联, 如创建“症状”类的实例“咳嗽 (cough)”, “症状”的数据属性“特征”与“病机”用作“咳嗽 (cough)”的属性声明, 其中“咳嗽 (cough)”个体有 6 条使用记录。创建“方剂”类的实例“止咳平喘剂”, “方剂”的数据属性“功用”用作“止咳平喘剂”的属性声明, 其中“止咳平喘剂”个体有 5 条使用记录, 其“功用”属性值为“有消除痰饮、止咳平喘及化痰、散结、截疟等作用”。中药主治症状, 而中药由方剂组成, 方剂与症状的语义

关系为“方剂 - 主治 - 症状”, 由上述语义关系可以构建二者实例关系, “咳嗽”的使用药物为“止咳平喘剂”, 对应“止咳平喘剂”可以主治“咳嗽”, 具体实例关系, 见图 7。

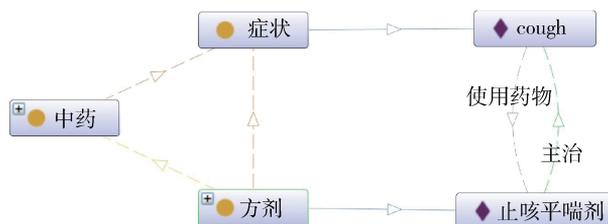


图7 咳嗽实例语义关系

3.4.2 实例2 “中药”可以影响“人体”, 为

具体展示中药对人体的影响关系，构建“人体”与“方剂”的实例关系。如创建“人体”类的实例“五脏”，“五脏”属于“脏腑”，“人体-脏腑”的数据属性“名称”与“生理功能”用作“五脏”的属性声明，其中“五脏”个体有 6 条使用记录。创建“方剂”类的实例“安神剂”，“方剂”的数据属性“方剂名”“用法”“功用”“配伍特点”用作“安神剂”的属性声明，其中“止咳平喘剂”个体有 8 条使用记录，其属性值如下。方剂名：安神剂；功用：安神药分别有镇惊安神、养血安神、补心安神及解郁安神等作用，此外有些安神药还有平肝潜阳、明目、收敛固涩等作用；用法：因神志不安常以虚实夹杂证出现，故重镇安神剂和滋养安神剂往往配合应用，以求标本兼顾；配伍特点：使用安神药需根据不同病因及病情变化适当配伍用药。如心火亢盛者，当配清泻心火药；痰火扰心者，当配清热化痰药等。中药影响人体，人体由脏腑组成，而中药由方剂组成，方剂与人体的语义关系为“方剂-影响-人体”，由上述语义关系可以构建二者实例关系，“五脏”的使用药物为“安神剂”，对应“安神剂”可以影响“脏腑-五脏”，具体实例关系，见图 8。

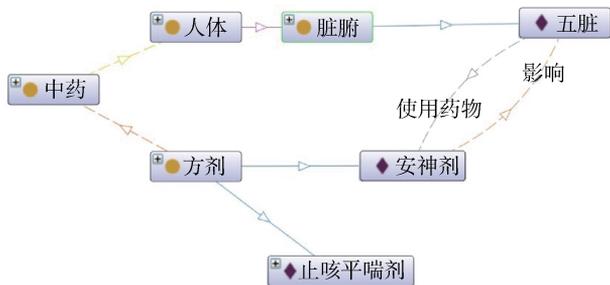


图 8 方剂与人体实例语义关系

4 创新性及其局限性

4.1 创新性

借鉴药典知识与 TCMLS Semantic Network 顶层本体框架，针对中医药学本体构建，初步建立中医药学本体框架，中药名称概念包含中药材、中药饮片和方剂，因各自内涵属性不同，将其概念分开描

述分析，以确保其类与类之间的概念独立性和准确性^[10]，使其体系更加完善，同时提高中医药知识表示的复用可能性。不同于以往构建中医本体的单一性，将方剂和中药材合并为中药，简化本体构建。本文创新点在于将方剂、中成药及中药材饮片单独分类构建，概念更为具体详尽，更有层次性；同时构建方剂与症状之间的实例，解释药物对症状的作用，具体阐述药理方面的解释，使对中药本体的理解更加深入。

4.2 局限性

在构建中医药本体的过程中存在一定局限性：一是中医药领域本体构建在不同研究中类的划分存在差异性，无法较好地重复利用；二是中药本体的对象属性构建在不同研究中无法完全统一，不同中药书籍对药品描述标准不同，会产生语义层面不统一，如何实现并制定统一的适配语义系统日益迫切；三是中药属性定义具有一定主观性，现有研究通常通过专家评价方式来解决该问题。建议将复杂概念层次化，建立较为完善统一的本体体系，以增加可利用性；继续完善和发展中医药学语言系统；参考已有的评价标准，建立评价体系指导本体构建；利用已有本体框架进行进一步知识推理，设计相应方剂应用场景，解决中医药针对不同病因病机的用药方案选择问题。

5 结语

本体构建有助于中医药领域知识表达和应用，中医药领域本体研究仍存在可探索的空间，为提高中医临床医疗信息利用效率，需加快完善中医药语言系统^[14]。如何将中医药信息知识全面展示的同时提高中医药领域本体实际应用的服务性还需探讨。

参考文献

1 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.

(下转第 47 页)

参考文献

- 1 董尔丹, 胡海, 俞文华. 生物样本库是生物医学研究的重要基础 [J]. 中国科学: 生命科学, 2015, 45 (4): 359-370.
- 2 满秋红, 王伟业. 重构生物样本库建设理念, 促进临床研究资源协同应用与发展 [J]. 协和医学杂志, 2019, 10 (4): 409-413.
- 3 崔晓英, 宋双, 翟倩雯, 等. 精准医学时代下中医特色生物资源样本库的建设探析 [J]. 江苏科技信息, 2020, 37 (9): 25-27, 40.
- 4 陈万思, 白雨菲. 新中国成立 70 年来劳动关系研究知识图谱与演进——基于 1949-2018 年间 5820 篇文献的计量分析 [J]. 中国劳动关系学院学报, 2020, 34 (3): 51-62, 84.
- 5 石晶晶, 石树青, 胡元会, 等. 心血管疾病能量代谢研究的可视化分析 [J]. 中国循证心血管医学杂志, 2020, 12 (4): 428-432, 442.
- 6 钱文静, 张有奎. 人类命运共同体研究的可视化知识图谱分析 [J]. 西南民族大学学报 (人文社科版), 2020, 41 (7): 222-230.

- 7 李海燕, 倪明宇, 林金嬉, 等. 临床生物样本库应用机制探索 [J]. 中华医学科研管理杂志, 2019, 32 (5): 397-400.
- 8 韩江, 周为民, 陈涵, 等. 如何发挥生物样本库在临床研究中的作用 [J]. 转化医学杂志, 2018, 7 (6): 346-349.
- 9 陈思静, 吴茂锋, 李佩娟. 生物样本库的建设与发展 [J]. 生物化工, 2019, 5 (4): 164-166.
- 10 赵励彦, 范贞, 刘瑞爽, 等. 生物样本库的伦理审查 [J]. 中国医学伦理学, 2020, 33 (3): 345-348.
- 11 雷瑞鹏, 冯君妍. 生物样本数据库的泡沫已经破裂了吗? ——数字化时代对生物样本数据库可持续发展的挑战 [J]. 医学与哲学, 2018, 39 (9): 90-95.
- 12 满秋红, 于农, 闫飞, 等. 我国生物样本库建设现状与未来发展的思考 [J]. 中国医药生物技术, 2018, 13 (4): 289-293.
- 13 刘峙雅. 我国生物样本库的研究进展 [J]. 现代医药卫生, 2021, 37 (5): 759-763.
- 14 郜恒骏, 杜莉利, 张小燕, 等. 生物样本库发展的现状、机遇与挑战 [J]. 协和医学杂志, 2018, 9 (2): 172-176.

(上接第 42 页)

- 2 Zhou X, Wu Z, Yin A, et al. Ontology Development for Unified Traditional Chinese Medical Language System [J]. Artificial Intelligence in Medicine, 2004, 32 (1): 15-27.
- 3 陈欢欢. 图书情报学领域本体的构建研究 [J]. 图书馆学研究, 2011 (11): 11-16.
- 4 郝伟学. 中医健康知识图谱的构建研究 [D]. 北京: 北京交通大学, 2017.
- 5 贾李蓉, 刘静, 于彤, 等. 中医药知识图谱构建 [J]. 医学信息学杂志, 2015, 36 (8): 51-53.
- 6 刘丽红, 刘静, 李海燕, 等. 《中国中医药学主题词表》“中药功能分类”主题词选词研究 [J]. 中国中医药信息杂志, 2015, 22 (10): 38-40.
- 7 肖洁, 冒宇清, 李新霞. 中医病证本体库的构建研究 [J]. 无线互联科技, 2018, 15 (13): 115-116, 133.
- 8 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 (6): 807-809.
- 9 侯玉, 张昌林, 周强, 等. Protégé 在构建中医治则治法

- 本体中的运用 [J]. 医学信息 (上旬刊), 2007, 20 (11): 1866-1871.
- 10 倪萍, 陆宇宏. 中医方剂领域本体建模和语义推理实现 [J]. 现代情报, 2012 (6): 133-140.
- 11 Miao F, Liu H, Huang Y, et al. Construction of Semantic-based Traditional Chinese Medicine Prescription Knowledge Graph [C]. Chongqing: 2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), 2018.
- 12 刘丽红, 贾李蓉, 刘静, 等. 中医临床术语系统 v2.0 “中药饮片”类概念及关系设定研究 [J]. 中国中医药信息杂志, 2019, 26 (1): 95-98.
- 13 国家药典委员会. 中华人民共和国药典: 2020 版 [M]. 北京: 中国医药科技出版社, 2020.
- 14 朱彦, 贾李蓉, 高博, 等. 中医临床术语系统 v2.0 设计与构建 [J]. 中国中医药图书情报杂志, 2018, 42 (3): 10-15.