

《4 种重大疾病的中医古籍叙词表》编制研究^{*}

付 兴 李芊芊 杨 凤 周冉冉 侯鉴宸 陶晓华

(北京中医药大学 北京 102488)

〔摘要〕 运用中医古籍“病脉证并治”知识元标引体系,结合现代文献研究,制定词汇过滤策略、选词范围、选词原则等,构建 4 种重大疾病(脑卒中、心力衰竭、肝硬化和糖尿病)的中医古籍叙词表树形结构,为后续 4 种重大疾病的中医临床术语知识框架构建奠定基础。

〔关键词〕 中医古籍;重大疾病;叙词表;知识元标引

〔中图分类号〕 R-058 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2022.12.008

Study on the Compilation of Glossary of Subject Headings of Ancient Chinese Medicine Books on Four Major Diseases FU Xing, LI Qianqian, YANG Feng, ZHOU Ranran, HOU Jianchen, TAO Xiaohua, Beijing University of Chinese Medicine, Beijing 102488, China

〔Abstract〕 By applying the knowledge meta-indexing system of “disease, pulse, evidence, and treatment” of ancient Chinese medicine books and combining with modern literature research, the paper formulates vocabulary filtering strategies, word selection scope and word selection principles, etc., and constructs the tree structure of glossary of subject headings of ancient Chinese medicine books for four major diseases (stroke, heart failure, liver cirrhosis and diabetes), so as to lay the foundation for the construction of the knowledge framework of Traditional Chinese Medicine (TCM) clinical terms for the four major diseases.

〔Keywords〕 ancient Chinese medicine books; major diseases; glossary of subject headings; knowledge meta-indexing

1 引言

中医古籍记载了我国数千年来医疗实践的经验,是传承中医科学和医疗创新的源泉和根基。但中医古籍数量众多、门类庞杂,要充分挖掘中医古籍的价值,需要深化基于知识元理论的中医古籍整理挖掘技术与方法。叙词表也称主题词表,以更加规范化的系统语言代替标引人员使用的自然语言,是中医古籍文献检索中进行文献主题标引的检索工具之一,是中医古籍知识库的基石,也是连接中医各种专业词语与计算机系统语言的纽带。但是,现有词表无法满足现代

〔修回日期〕 2022-05-05

〔作者简介〕 付兴,博士,发表论文 10 篇;通信作者:陶晓华,教授,博士生导师,发表论文 88 篇。

〔基金项目〕 国家重点研发计划“基于知识元理论与临床需求深度融合的中医古籍整理及专题文献研究”(项目编号:2019YFC1709200);国家重点研发计划课题“基于知识元理论的中医古籍整理挖掘技术与方法研究”(项目编号:2019YFC1709201)

疾病与中医病证名的相关性研究需求。因此,本研究基于 600 种中医古籍的深度标引工作,选取 4 种中医诊疗具有一定优势和特色的疾病——脑卒中、心力衰竭、肝硬化和糖尿病,通过叙词表构建,将相关中医词汇进行集合,用以指导后续的标引和检索,揭示概念间关系,也可为后续 4 种重大疾病中医临床术语知识框架建立奠定基础。

2 医学叙词表研究现状

2.1 叙词表

叙词表是概括各门或某一学科领域并由语义相关、族性相关的术语所组成的、可以不断补充的规范化词表^[1]。叙词表主要由叙词与叙词间关系组成,叙词间基本语义关系采用不同的参照符号以示区分。中国《汉语主题词表》采用“用、代、属、分、参”结构。一般叙词表中包括等同关系、等级关系和相关关系 3 种关系^[2]。叙词表作为一种术语控制工具,可以通过将古籍资料、文献、检索者的自然语言等转换成相对规范化的语言,从而形成某特殊领域内的表达事物概念的词汇集合;还可以通过各方式对各叙词之间的各种语义联系进行显示,是可以不断补充、完善和调整的动态词表^[3]。

2.2 现代医学叙词表研究现状

国外主要的医学叙词表包括美国国家癌症研究所的《NCI 主题词表》(NCIt)和美国国家医学图书馆(National Library of Medicine, NLM)编制的《医学主题词表》(MeSH)等,其中 MeSH 影响力最大^[4]。MeSH 最早出版于 1960 年,经过医学学科专家和 MeSH 工作人员的不断修订和更新,2019 版 MeSH 收录了 29 351 个概念和 247 209 个补充概念记录。这些补充概念来源于生物医学领域的其他相关词表。MeSH 具有独特的主-副主题词组配表达形式,采用树状结构表代替传统的分类表。MeSH 用于标引世界领先的生物医学数据库 PubMed 的文献,也用于 NLM 编目数据库的书目描述。

2.3 中医药学叙词表研究现状

2.3.1 中医药学主要叙词表 现代中医药词表的

研制工作起步于 20 世纪 70 年代^[5]。我国现有的医学领域叙词表包括医学(现代医学)、药学和中医药 3 个领域,代表性叙词表有:中国医学科学院医学信息研究所将 MeSH 翻译为中文,并在此基础上建立的“中文医学主题词表检索系统”(CMeSH)^[6],是国内医学领域权威的叙词表;国家食品药品监督管理局于 2013 年编制出版的《中国药理学主题词表》,共收录术语 29 000 条^[7];中国中医科学院中医药信息研究所编制的《中国中医药学主题词表》,1987 年出版第 1 版,2015 年发布网络版,目前共收录术语 21 466 条^[8]。其中,“中文医学主题词表检索系统”和《中国中医药学主题词表》都建立了持续更新机制,而且二者还建立了网络版服务系统^[9]。此外,中国中医科学院从 2002 年开始,借鉴统一一体化语言的经验,研制了“中医药学语言系统(Traditional Chinese Medicine Language System, TCMLS)”,对中医药学的名词术语进行了系统梳理,构建了中医药概念术语的层次结构和复杂语义网络。已收录约 12 万个概念、30 万个术语以及 127 万条语义关系^[10-11]。定位在专门面向中医临床的大型术语系统“中医临床术语集”,已收录 11 万多条概念词,27 万多个术语,内容覆盖临床所见、病证、操作、治则治法和中药等中医临床知识各个方面^[12]。建立了概念与概念之间的层级关系、概念与概念之间的相关关系。

2.3.2 问题分析 目前基于本体技术的中医药知识库系统成为中医药信息化领域的主流研究方向^[13]。叙词表采用一种树形结构方式展现中医古籍中的概念结构关系,表现出来自古籍的内容与概念,是对中医古籍文献的客观反映,也展现了古籍文献本身的结构。领域本体的编制以叙词表为依据,继承了叙词表的树状结构特征^[14]。然而,尚缺乏以现代疾病为研究对象且基于中医古籍词语的叙词表,尽管中西医是两种不同的学科体系,但二者的研究目的与对象是一致的。因此,为架设传统医学与现代医学的研究桥梁,构建具有中医特色的《4 种重大疾病中医古籍叙词表》,具有重要的研究意义与临床价值。

3 《4 种重大疾病的中医古籍叙词表》构建原则

3.1 编表原则与标准

为满足中医药领域内 4 种重大疾病的研究工作需求,采用“以主题为主,主题与学科相结合”的原则,突出叙词表中 4 种重大疾病的专业性,保证实用性,充分考虑地域和时代。词表的编制标准参考《汉语叙词表编制规则 (GB 13190—1991)》和《军用主题词表编制规则 (GJB1776A—99)》制定。语料构建参考《中国药学主题词表》《中国中医药学主题词表》《健康信息学 - 中医药学语言系统语义网络框架 (GB/T 38324—2019)》以及中医语言系统发布平台等。

3.2 选词原则

词表选词遵循以下 3 个原则,一是保证参考标引频率,选用在中医药领域较常出现,具有一定使用频率的名词术语;二是保证基本词汇准确,参考《中国中医药学主题词表》的基础上,保证与 MeSH 词表内中医药相关词语兼容,同时以《国际疾病分类第十一次修订本 (ICD11)》《中医病证分类与代码 (GB/T 15657—2021)》的术语内容为标准;三是保证所选词语概念清晰简练,词义明确。

4 《4 种重大疾病的中医古籍叙词表》构建

4.1 中医古籍标引平台与模板

本研究基于采用“病脉证并治”知识元标引技术与方法的中医古籍标引平台 (<http://114.255.40.130:60080/metaservice2/#>),在对中医古籍进行标引时,以“病脉证并治”标引模版为

首选,并根据不同类别、不同内容的古籍配置 9 种标引模版及元数据。其中“病脉证并治”标引模版以中医疾病名为核心,包括病名、病因、病机、治则、治法、主脉、兼脉、辨证、方剂、药物、针灸、预后、宜忌等 30 个知识元,以及 59 种语义类型。9 种标引模版及元数据包括中医理论、诊法、病证、本草、方剂、医案、预防调护、学术流派、针灸标引模版。

4.2 4 种重大疾病相关中医古籍语料构成

现代疾病与中医病名的对应问题一直是学界争议的关键。首先,本研究以 4 种重大疾病及其并发症为对象,以现代文献研究为抓手,以 4 种重大疾病相关的中医疾病名为核心,在中国知网文献数据库中以脑卒中、心力衰竭、肝硬化、糖尿病、并发症及其同义词为关键词进行精确检索,共检索 1995 年 1 月 1 日—2022 年 2 月 1 日的中文文献 6 003 篇,其中脑卒中 519 篇,心力衰竭 445 篇,肝硬化 947 篇,糖尿病 4 092 篇,提取文中涉及的并发症名称,将相同病名进行合并,如中风后抑郁、卒中后抑郁、脑卒中抑郁症合并为抑郁症,从中归纳出 4 种重大疾病及其并发症共 208 种,其中脑卒中 76 种,心力衰竭 15 种,肝硬化 45 种,糖尿病 72 种;其次,再以筛选出的 4 种重大疾病及其并发症名称、中医病名或中医为关键词,在中国知网文献数据库中再次进行精确检索,去除缺乏中医认识的部分现代疾病,例如电解质紊乱、酸碱失衡等;最后,参考《国际疾病分类第十一次修订本 (ICD11)》和 2020 年颁布的《中医临床诊疗术语第 1 部分:疾病》以及相关临床指南与专著等,剔除部分欠规范的中医病名,最终形成 4 种重大疾病及其并发症相关的中医疾病名 318 种,其示例,见表 1。

表 1 4 种重大疾病并发症中西医病名对应

重大疾病	并发症 - 西医病名	并发症 - 中医病名
脑卒中	肠炎 ^[15]	肠澼;便血;赤沃;飧泻;泄泻;大肠泄;小肠泄;大瘕泄;下利;滞下;痢;休息痢;久痢;痢疾;脏毒 ^[16]
	肺炎 ^[17]	痰饮;咳嗽;肺炎喘嗽病;肺热病;喘病;肺痿病;咯血病;感冒病;肺胀病;风温病;风湿肺热病;外感高热病 ^[18]
.....		

基于中医古籍“病脉证并治”知识元标引体系，以中医疾病名为导向，突显“以病为纲”（由于 4 种重大疾病的并发症存在交叉，而其并发症的中医病名并无区别，因此未对其并发症的中医病名进行区分），同时参考《中国中医药学主题词表》等现有叙词表，制定现代医学病名、中医疾病名、中医理论、病因病机、诊法、症状体征、推荐方剂、中药、治疗、预防调护、古籍书目、学术流派及医家共 12 个范畴。以 4 种重大疾病及其并发症相关的中医疾病名遴选知识体，将其中的知识元与语

义提取出来，将不同类型语义归入不同范畴类目中，如语义类型为治法的语义“滋阴”归入“治疗”范畴下的“治法”中，从而形成 4 种重大疾病相关的中医古籍叙词表语料库。

4.3 《4 种重大疾病的中医古籍叙词表》构建流程

构建中医古籍叙词表流程包括选词过程、编制主表、编制词族表、编制范畴表、编制英汉索引表等内容，见图 1。

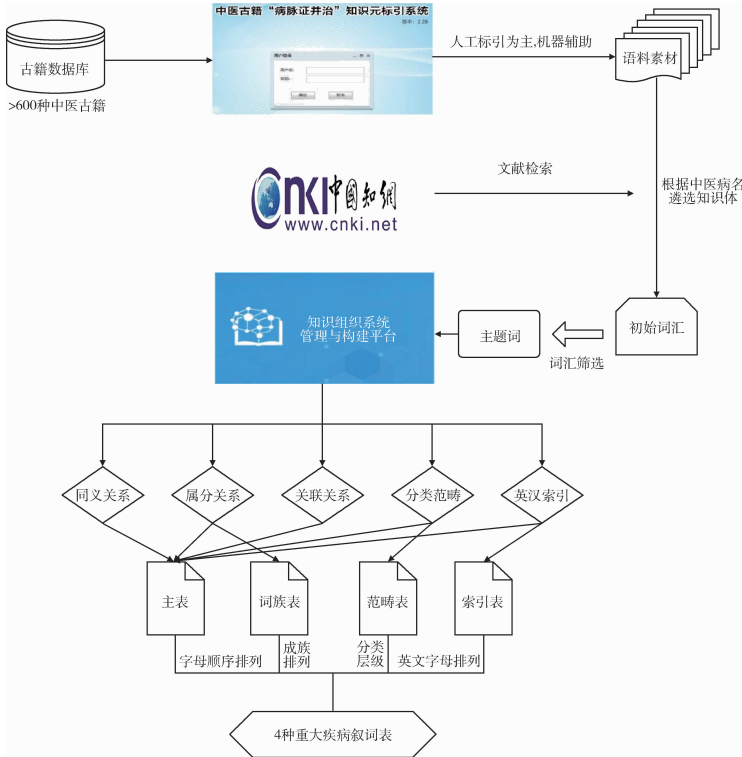


图 1 《4 种重大疾病的中医古籍叙词表》构建流程

4.3.1 选词过程 对中医古籍进行原始版本数字化处理，形成计算机可读文本文件，并按“病脉证并治”知识元核心要素进行深度标引加工，遵循“一标、二查、三复审、四抽检”的原则，对标引后的条目进行人工和机器“病脉证并治”知识元核心要素双分析，利用自然语言处理等技术手段，对数字化后的古籍原文内容进行机器分析挖掘，提取出其中的词汇及词汇出现的频次，构建 4 种重大疾病中医古籍叙词表的基本素材词库。采用自然语言处理等技术提取词汇（尤其是高频词汇）。参考各

类文献数据库，以及中医叙词表与词典工具书等传统纸质文献和书籍资料，如《中国中医药学主题词表》《中医大辞典》等，并保证与 MeSH 词表内中医药相关词语兼容。

4.3.2 编制主表 通过以上选词来源，基于网络关键字统计系统统计期刊的关键字，在分词结果和关键字统计结果中，通过制定词频阈值、词汇过滤规则、选词范围、选词原则来选取正式主题词。

4.3.3 编制词族表 采用本课题合作单位搭建的“知识组织系统管理与构建平台”构建词汇语义成

族的结构体系，其中包括展示词汇的范畴号、英汉对照关系、同义关系、等级关系、关联关系等内容。在编制主表过程中，制定正式主题词与非正式主题词的编制结构体系。在编制词族表过程中，制定词汇语义成族的结构体系，属、分、族项通过导入词族表的等级关系来实现，范畴号由导入范畴表中设定的类别号来实现。

4.3.4 编制范畴表、编制英汉索引表 制定范畴表编排的结构体系，制定汉语拼音索引、英汉对照索引的结构体系。本叙词表研究旨在中医古籍知识分类体系下进行扩展与深化，尤其关注以疾病为纲的中医古籍叙词编著与研究。建立叙词的概念关系，借鉴词表概念间的逻辑关系，结合古籍知识库语义元数据，参考中医药一体化语言系统，采用自下而上与自上而下相结合的方法，构建中医疾病古籍叙词表的概念语义网络。

4.4 《4 种重大疾病的中医古籍叙词表》基本内容

《4 种重大疾病的中医古籍叙词表》包括主表、范畴表、词族表和索引表（汉语拼音索引、英汉对照索引）。其中范畴表类目，见表2。该词表涵盖现代医学病名、中医疾病名、中医理论、病因病机、诊法、症状体征、推荐方剂、中药、治疗、预防调护、古籍书目、学术流派及医家共12 个范畴，在各类目之下，再按中医理论逐项细分，最深达7 级，见表3。目前收录包含树形号、中文、英文、拼音、定义或简介以及范畴号的正式主题词7 234 个，非正式主题词5 255 个。语义关系参考《健康信息学－中医药学语言系统语义网络框架（GB/T 38324—2019）》，包括上下位关系、相关关系、分支、包含等10 余种。

表2 《4 种重大疾病的中医古籍叙词表》范畴表类目

范畴号	类目	范畴号	类目	范畴号	类目	范畴号	类目
TA	现代医学疾病名	TA1	脑卒中	TA2	慢性心衰	TA3	糖尿病
TA4	肝硬化	TB	中医疾病名	TB1	中风病	TB2	心病病
TB3	消渴病	TB4	积聚类病	TB5	并发症	TC	中医理论
TC1	五行理论	TC2	气血津液精	TC3	经络腧穴	TC4	营卫
TC5	藏象	TC6	阴阳学说	TD	诊法	TD1	切诊
TD2	望诊	TD3	问诊	TD4	闻诊	TE	病因病机
TE1	病因	TE2	病机	TF	症状体征	TF1	神志症状
TF2	脉象表现	TF3	身体形态	TF4	姿态	TF5	皮肤表现
TF6	舌苔表现	TF7	睡眠症状	TF8	情志症状	TF9	言语表现
TF10	毛发表现	TF11	四肢症状	TF12	二便症状	TF13	呼吸症状
TF14	饮食口味症状	TF15	一般症状	TF16	躯体症状	TF17	颈项症状
TF18	五官症状	TF19	头面症状	TG	治疗	TG1	治则类术语
TG2	治法类术语	TG3	疗法类术语	TH	推荐方剂	TH1	脑卒中推荐方剂
TH2	慢性心衰推荐方剂	TH3	糖尿病推荐方剂	TH4	肝硬化推荐方剂	TI	中药（功效分类）
TI1	安神药	TI2	拔毒化腐生肌药	TI3	补益药	TI4	涌吐药
TI5	攻毒杀虫止痒药	TI6	化湿药	TI7	化痰止咳平喘药	TI8	解表药
TI9	开窍药	TI10	理气药	TI11	理血药	TI12	利水渗湿药
TI13	平肝熄风药	TI14	清热药	TI15	驱虫药	TI16	祛风湿药
TI17	祛风药	TI18	祛暑药	TI19	软坚药	TI20	收涩药
TI21	外用药	TI22	温里药	TI23	消食药	TI24	泻下药
TI25	止痛药	TJ	预防调护	TJ1	预防	TJ2	护理
TJ3	康复	TK	古籍书目	TK1	本草类	TK2	儿科类
TK3	方书类	TK4	妇科类	TK5	其他类	TK6	金匱类

续表 2

TK7	内科类	TK8	伤寒类	TK9	外科类	TK10	温病类
TK11	养生类	TK12	医案类	TK13	医话类	TK14	医经类
TK15	综合类	TK16	针灸类	TK17	诊法类	TL	学术流派及医家
TL1	伤寒学派	TL2	医经学派	TL3	温病学派	TL4	汇通学派
TL5	针灸学派	TL6	本草学派	TL7	易水学派	TL8	河间学派
TL9	丹溪学派	TL10	新安医派	TL11	孟河医派	TL12	钱塘医派
TL13	永嘉医派						

注：中药根据功效分类进行分类。

表 3 范畴各级数目（个）

范畴名称	2 级	3 级	4 级	5 级	6 级	7 级	合计
现代医学病名	4	0	0	0	0	0	4
中医疾病名	5	51	36	123	69	33	317
中医理论	6	22	82	42	0	0	152
诊法	4	16	48	33	9	0	110
病因病机	2	14	59	143	95	0	313
症状体征	18	269	446	177	134	37	1 081
治疗	3	45	274	428	352	66	1 168
推荐方剂	4	2 058	0	0	0	0	2 062
中药（功效分类）	25	173	803	122	0	0	1 123
预防调护	3	10	6	0	0	0	19
古籍书目	17	618	0	0	0	0	635
学术流派及医家	16	234	0	0	0	0	250

5 结语

综上所述，基于中医古籍的 4 种重大疾病叙词表旨在立足于中医古籍知识分类体系下进行扩展与深化，探索性试以现代文献研究为依据，寻求现代医学病名与中医疾病名的连接。采用具有“病脉证并治”特色的中医古籍知识元标引方法，完善以中医疾病名为纲的中医古籍叙词编著与研究，从而形成 4 种重大疾病的中医古籍叙词表。构建 4 种重大疾病的中医古籍叙词表后，为其他现代疾病的叙词表建立提供范本与参考，或可为构建 4 种重大疾病的中医临床术语知识框架奠定良好基础。

参考文献

1 曾雅萍，张魁清，邱祖斌．航空科技信息检索语言探索－实践－研究五十年 [C]．北京：第四届全国知识组织与知识链接学术交流会，2013.

2 杨继红．基于本体的中医古籍叙词表构建方法研究 [D]．北京：中国中医科学院，2008.

3 杨建伟，张志平．叙词表的 OWL 表示方法研究 [J]．情报杂志，2007（3）：19－21.

4 彭莉，储戟农，范为宇，等．MeSH 词表的管理 [J]．国际中医中药杂志，2013，35（12）：1107－1109.

5 张晶，范为宇，彭莉，等．中医药学及其相关主题词表概述 [J]．国际中医中药杂志，2011（5）：436－439.

（下转第 56 页）

- 构建、推理与知识超图理论 [J]. 计算机应用, 2021, 41 (8): 2161–2186.
- 6 中国中文信息学会. 全国知识图谱与语义计算大会 (CCKS 2019) 技术评测任务发布 [EB/OL]. [2021–11–19]. <http://www.cipsc.org.cn/chcontent.php?xuhao=2018031804>.
- 7 Xie X, Zou Y F, Zhang S W, et al. Research on the Construction of Knowledge Map of Chinese Film and Television [J]. Journal of the University of South China (Natural Sciences), 2019, 33 (1): 73–79.
- 8 Zhai S P, Guo L, Gao S, et al. Method for Knowledge Graph Completion Based on Bayesian Reasoning [J]. Journal of Chinese Computer Systems, 2018, 39 (5): 995–999.
- 9 Devlin J, Chang M W, Lee K, et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding [EB/OL]. [2021–11–19]. <https://arxiv.org/pdf/1810.04805.pdf>.
- 10 Hochreiter S, Schmidhuber J. Long Short-term Memory [J]. Neural Computation, 1997, 9 (8): 1735–1780.
- 11 Lafferty J, McCallum A, Pereira F C N. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data [EB/OL]. [2021–11–19]. https://repository.upenn.edu/cgi/viewcontent.cgi?article=1162&context=cis_papers.
- 12 Lan Z, Chen M, Goodman S, et al. ALBERT: a Lite BERT for Self-supervised Learning of Language Representations [EB/OL]. [2022–05–10]. <https://arxiv.org/pdf/1909.11942.pdf>.
- 13 Sun Z, Li X, Sun X, et al. ChineseBERT: Chinese Pre-training Enhanced by Glyph and Pinyin Information [EB/OL]. [2021–11–19]. <https://arxiv.org/pdf/2106.16038.pdf>.
- 14 刘千军, 桂前进, 张祺, 等. 实体关系抽取技术研究进展综述 [J]. 计算机应用研究, 2020, 37 (S2): 1–5.
- 15 杨穗珠, 刘艳霞, 张凯文, 等. 远程监督关系抽取综述 [J]. 计算机学报, 2021, 44 (8): 1636–1660.
- 16 郑德权, 胡佳. 远程监督关系抽取方法综述 [J]. 计算机应用, 2021, 41 (S1): 7–14.

(上接第 44 页)

- 6 李丹亚, 胡铁军, 李军莲, 等. 中文一体化医学语言系统的构建与应用 [J]. 情报杂志, 2011, 30 (2): 147–151.
- 7 储戟农, 彭莉, 崔蒙, 等. 论《中国中医药学主题词表》在中医药信息学中的作用 [J]. 国际中医中药杂志, 2014, 36 (12): 1060–1063.
- 8 阮学平. 关于医学分类表与主题词表再对应的思考 [J]. 医学情报工作, 2002 (3): 181–183.
- 9 张晶. 《中国中医药学主题词表》用户使用需求调查分析 [C]. 北京: 中国中医科学院中医药信息研究所 2011 年学术年会, 2012.
- 10 尹爱宁, 张汝恩. 建立中医药学一体化语言系统的方法学研究 [J]. 中国中医药信息杂志, 2003 (9): 90–92.
- 11 贾李蓉, 刘丽红. 基于中医药学语言系统的文献检索服务平台 [J]. 医学信息学杂志, 2012, 33 (1): 54–56.
- 12 杨阳, 李园白, 崔蒙. 建立中医临床术语集探索性研究 [J]. 中国中医药信息杂志, 2006 (12): 105–110.
- 13 于彤, 李敬华, 于琦. 中医药知识服务研究概述 [C]. 北京: 第一届中国中医药信息大会, 2014.
- 14 于彤, 杨硕, 李敬华. 中医药知识库系统的研究进展 [J]. 中国医学创新, 2014, 11 (18): 142–144.
- 15 龙凯, 李雄文. 早期肠内营养支持对重症脑卒中患者营养状况及并发症的影响 [J]. 实用心脑血管病杂志, 2014, 22 (9): 114–115.
- 16 甄建华, 黄光瑞. 溃疡性结肠炎中医病名、病因、病机的古今比较和回顾 [J]. 环球中医药, 2019, 12 (8): 1286–1289.
- 17 胡徽星, 梁发俊, 张静波, 等. 俞募配穴针刺联合益气化痰方治疗卒中后相关性肺炎 40 例 [J]. 环球中医药, 2021, 14 (9): 1705–1708.
- 18 王见义, 竺丽明, 陈建新, 等. 基于熵聚堆方法对肺炎中西医病名标准的对照研究 [J]. 中医杂志, 2010, 51 (S2): 165–166.