

# 我国临床医学术语体系建设路径与实践探索\*

任慧玲 李亚子 李晓瑛 冀玉静 邓盼盼

(中国医学科学院/北京协和医学院医学信息研究所 北京 100020)

**[摘要]** 结合医学术语标准理论与实践研究经验, 阐述中文临床医学术语体系建设路径, 包括顶层框架设计、底层内容建设、关键技术平台研发等, 有助于实现国内临床医学信息的规范表示、互通共享与挖掘利用, 并对我国乃至其他母语非英语国家的医学术语标准化提供一些借鉴。

**[关键词]** 临床医学术语; 术语标准化; 建设路径

**[中图分类号]** R-058 **[文献标识码]** A **[DOI]** 10.3969/j.issn.1673-6036.2022.08.001

**The Construction Path and Practical Exploration of Clinical Medical Terminology System in China** REN Huiling, LI Yazi, LI Xiaoying, JI Yujing, DENG Panpan, Institute of Medical Information, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100020, China

**[Abstract]** Based on the theoretical and practical research experience of medical terminology standards, the paper expounds the construction path of Chinese clinical medical terminology system, including top-level framework design, underlying content construction, key technology platform development, etc. The research findings and results are helpful to realize the standardized representation, sharing, mining and utilization of clinical medical information in China, and provide some references for the standardization of medical terminology in China and even other non-English speaking countries.

**[Keywords]** clinical medical terminology; standardization of terminology; construction path

## 1 引言

《“健康中国2030”规划纲要》明确指出, “共建共享、全民健康”是建设健康中国的基本路径与

根本目的。术语指用于命名特定领域概念的名称, 医学语言丰富, 临床数据种类多, 产生了大量临床医学术语。临床医学术语体系对临床常用术语及其相互关系进行描述与组织。建设临床医学术语体系的主要目标是对临床诊疗活动中所使用的基本语言进行规范命名与统一界定, 以解决临床医学术语命名重复、内涵外延不清晰、文字表达和语义理解不一致等问题, 促进健康医疗大数据互联互通、共建共享以及电子病历等现实世界数据的再次科研利用, 加快医疗卫生信息标准化进程与“健康中国”战略落地实施。

《系统化临床医学术语集》(Systematized No-

**[修回日期]** 2022-08-18

**[作者简介]** 任慧玲, 研究馆员, 发表论文50余篇; 通信作者: 李晓瑛, 博士, 副研究员, 发表论文40余篇。

**[基金项目]** 科技创新2030——“新一代人工智能”重大专项课题“中文医学术语体系构建”(项目编号: 2019AAA0104901)。

menclature of Medicine Clinical Terms, SNOMED CT) 等国外产权的临床医学术语体系规模体量庞大、语义关系丰富、更新维护持久、全球应用广泛<sup>[1-2]</sup>, 已成为国际主流标准。我国主要采取引进翻译、自主审定名词与融合创新构建中医药学术语标准等方式, 编制了《国际疾病分类法》第10次修订本(International Classification of Diseases V10, ICD-10)中文版、《常用临床医学名词》和《中国中医药学主题词表》等; 但各中文医学术语标准之间相对独立, 缺乏映射关联, 尚未形成能够覆盖临床医学全主题、支撑健康医疗大数据联通交互与分析利用的大规模、权威性体系。

中国医学科学院医学信息研究所在40余年医学术语标准理论与实践研究的基础上, 逐渐探索出一条借鉴融合与自主创新相结合的中文临床医学术语体系建设之路, 其中涉及顶层框架设计、底层内容建设、关键技术平台研发、与国际临床医学术语体系映射互操作等重要环节。目前初步建成具有自主知识产权的中文临床医学术语体系, 覆盖14个维度、20余万概念、60余万术语及上百种语义关系类型。本研究对中文临床医学术语体系建设路径与应用实践进行探索, 以期为我国及其他非英语国家的医疗卫生信息标准化建设提供借鉴。

## 2 国内外发展现状

### 2.1 国外

2.1.1 总体情况 国际卫生机构、欧美等发达国家十分重视临床医学术语体系标准化建设。世界卫生组织(World Health Organization, WHO)、国际卫生术语标准开发组织(International Health Terminology Standards Development Organization, IHTSDO)、美国国立医学图书馆(U. S. National Library of Medicine, NLM)、美国雷根斯基夫研究院(U. S. Regenstein Institute)、国际人用药品注册技术协调会(International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use, ICH)等机构持续投入, 长期开展《国际疾病分类法》<sup>[3]</sup>、SNOMED CT<sup>[4]</sup>、《临床药物标准命名术语

表》(RxNorm)<sup>[5]</sup>、《观测指标标识符逻辑命名与编码系统》(Logical Observation Identifiers Names and Codes, LOINC)<sup>[6]</sup>、《监管活动医学词典》(Medical Dictionary for Regulatory Activities, MedDRA)<sup>[7]</sup>等术语标准编制与更新<sup>[8]</sup>, 在全球临床医学术语命名统一及健康医疗大数据互联互通中发挥积极作用<sup>[9]</sup>。

2.1.2 SNOMED CT 国际公认最庞大的临床医学术语集。最初由美国病理学会(U. S. College of American Pathologists, CAP)于1965年编著出版。历经50余年的发展, SNOMED CT层级体系从最初4个轴扩展为当前19个, 内容覆盖疾病、临床发现、身体结构、生物体、化学物质、物理对象、物理力、标本等临床医学信息; 2022年3月1日发布的SNOMED CT国际版包含361 780个概念、1 579 468条术语及278种语义关系类型、7 170 238条关系实例(绝大部分具有方向性), 并给出8 941条概念定义信息。从2007年4月起IHTSDO接手SNOMED CT维护、发布及知识产权等相关事宜。2022年成员包括43个国家和地区并已向3万余个人和组织颁发了会员执照。目前美国、英国、加拿大、新西兰和澳大利亚等国家已经指定SNOMED CT为临床信息系统的术语参考标准, 超过25个会员方可免费使用标准, 80多个国家和地区陆续开展了临床记录、知识表示、汇聚和分析等不同程度的应用。

### 2.2 国内

2.2.1 总体情况 近40年来我国政府机构、科研院所相继开展中文临床医学术语编制与推广实践。国家卫生健康委员会、全国科学技术名词审定委员会、中华医学会、中国中医科学院中医药信息研究所等机构深入开展中文医学术语标准研制及国际医学术语的翻译融合, 推动了我国医疗卫生信息化和医学科技创新发展。我国临床医学术语标准化工作主要分为3种模式: 引进翻译、自主审定名词与融合创新构建中医药学术语标准, 部分医疗健康信息化组织发布了自建的医学术语集(如浙江数字医疗卫生技术研究院 OMAHA 七巧板医学术语集)。尽管我国已在中文临床医学术语标准化工作中取得不少成就, 但仍然存在一定问题, 主要表现为各术语

标准之间相对独立、缺乏映射关联且尚未形成能够覆盖临床医学全学科主题、支撑健康医疗大数据联通共享与分析利用的中文临床医学术语体系。

2.2.2 引进翻译 现代医学源自西方，引进翻译国际医学术语标准是母语非英语国家常用的一种医疗卫生信息标准化模式。我国已完成疾病、手术操作、检查检验、药物不良反应等子领域的国际临床医学术语体系本地化改造，代表性成果包括 ICD、ICD - 9 - CM - 3、LOINC、MedDRA 的中文版。其中 ICD - 10 中文版是我国病案首页中疾病诊断的主要参考编码，由设置于北京协和医院的世界卫生组织疾病分类合作中心于 1996 - 1998 年编译完成。2001 年原卫生部要求从 2002 年 1 月 1 日起所有住院患者信息编码强制采用该标准<sup>[10]</sup>。中华医学会及有关医疗机构更新编译 ICD - 11 中文版，2018 年 12 月由国家卫生健康委员会印发，要求各地采用，未来将逐步投入临床应用。中国医疗信息化行业奠基人李包罗教授于 2003 年前后组织翻译了 SNOMED CT 早期版本，推动了国内医学信息学的发展壮大。

2.2.3 自主审定名词 国家卫生健康委员会、全国科学技术名词审定委员会在中文临床医学名词审定发布工作中做出突出贡献。2019 年底国家卫生健康委员会发布《常用临床医学名词》，主题覆盖内、外、妇、儿、肿瘤等 30 个临床专业，术语细分疾病诊断、症状体征、手术操作和临床检查 4 个类别，总数共计 42 000 余条，适用于我国各级各类医疗卫生机构临床病历及各类临床文书中相关内容的书写。从 1989 年至今全国科学技术名词审定委员会已陆续公布妇产科、口腔科、医学遗传学、心血管病学、外科、中医药学等权威医学名词分册 45 本，要求全国各科研、教学、生产经营以及新闻出版等单位遵照使用。

2.2.4 融合创新构建中医药学术语标准 中医药学术语标准是全世界传统医学信息标准化实践的典范。先后发布国家标准《中医病证分类与代码》(GB/T 15657 - 1995)、《中医临床诊疗术语》(GB/T 16751 - 1997)、《中医基础理论术语》(GB/T 20348 - 2006) 以及《中医基本名词术语中英对照国际标准》<sup>[11]</sup> 等术语标准，基本满足海内外中医药

学及相关行业的医疗、教学和科研人员的使用及交流需要。中国中医科学院中医药信息研究所参照《医学主题词表》(Medical Subject Headings, MeSH) 及 SNOMED CT 体系架构，编辑出版《中国中医药学主题词表》(Traditional Chinese Medicine Subject Headings, TCMESH) 及《中医临床术语系统》(Traditional Chinese Medicine Clinical Terminology System, TCMCTS)，在中医药学领域术语规范和数据库建设中发挥着重要支撑作用。

### 3 建设路径探索

中国医学科学院医学信息研究所长期深入开展中文医学术语标准相关理论研究与应用实践。面向中外文生物医学文献信息组织需要编制出版《中文医学主题词表》(Chinese Medical Subject Headings, CMeSH)、《中文一体化医学语言系统》(Chinese Unified Medical Language System, CUMLS)、《科技知识组织体系》(Scientific Technology Knowledge Organization System, STKOS) (医学部分)。2017 年起承担院校创新工程“中文临床医学术语系统构建研究”项目，2020 年起主持科技创新 2030——“新一代人工智能”重大专项课题“中文医学术语体系构建”，2021 年执行中国 - 世卫组织 2020 - 2021 双年度合作项目“中文临床医学术语框架体系及在呼吸系统疾病中的试点研究”，逐渐探索出一条借鉴融合与自主创新相结合的中文临床医学术语体系建设之路，其中涉及顶层框架设计、底层内容建设、关键技术平台研发、与国际临床医学术语体系映射互操作等重要步骤，见图 1。



图 1 中文临床医学术语体系建设路径

### 3.1 顶层框架设计

面向我国医疗健康和现实世界数据研究需求，制定中文临床医学术语体系顶层框架，内容涵盖常见疾病、解剖部位、临床表现、检查检验、手术操作、化学药品等领域重要概念；设计语义知识模型，规划概念间等级关系 (is\_a) 和表示临床信息的具体语义关系，用以解决临床用语同义、多义、表达不规范及语义缺失等问题；制定中文临床医学术语统一编码规范，便于永久管理及后续互操作映射。

### 3.2 底层内容建设

遵照顶层框架设计，沿着术语 - 概念 - 知识这一脉络，分专科完成底层内容建设。多渠道收集我国已有卫生信息标准与临床医学术语数据，以概念为单元完成同义术语汇聚，进而建立概念之间的语义知识关联。邀请临床、科研等多方专家参与内容审核，提高术语体系的科学性、实用性及完整性。中文临床医学术语体系建设的核心是大规模多源异构异质的中文临床术语采集、抽取、发现与整合。本研究提出自顶向下的多类型词表同义整合，以及自底而上的多源临床实体发现与标准化共两条不同的路线完成这一关键过程，见图 2。当前我国已经积累了多种类型的临床医学术语资源，包括国际临床医学术语表（如 ICD - 10、ICD - 9 - CM）的中文翻译及本地化改造版本、国家标准（如《疾病分类与代码》GB/T14396 - 2016）、卫生行业标准（如《结核病分类》WS196 - 2017）、劳动安全行业标准（如《社会保险药品分类与代码》LD/T90 - 2012）以及国家卫生健康委员会发布的《常用临床医学名词》《国家基本药物目录》等，具有极高的继承和发展价值；经一系列的清洗、去重、编码等数据预处理，这些资源中的临床医学术语可通过来源表颗粒度判别与控制<sup>[12]</sup>，完成对来源表已有同义关系的继承；进一步利用词频余弦相似度、Jaccard 系数等多种文本相似度算法，开展来自不同词表的术语同义计算，从而实现以概念为基本单元的同义术语汇聚；借助概念的定义、语义类型等属性及上下位语义关系，对同形异义的歧义术语添加特定标识以示

区别。

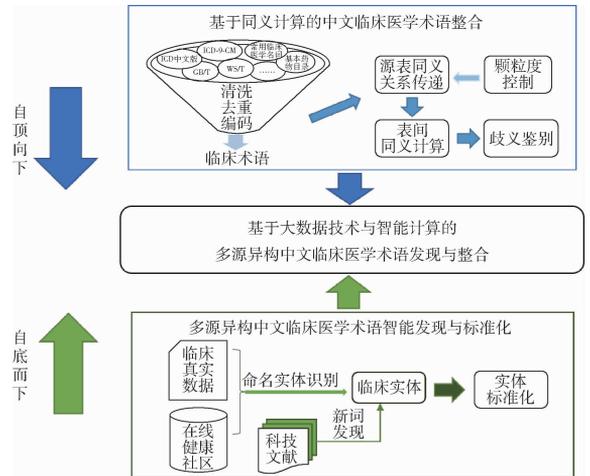


图 2 基于大数据技术与智能计算的多源异构中文临床医学术语发现与整合框架

### 3.3 关键技术平台研发

研究自动采集存储、术语同义整合、语义关系计算等术语体系构建的关键技术，开发多人协同加工审核平台及术语检索展示平台，实现计算机辅助人工的大规模术语体系建设模式，提升建设效率与内容一致性。互联网信息时代背景下，临床医学术语的另一个主要来源是健康医疗大数据，包括电子病历、个人健康档案等临床真实世界数据，医患之间便捷交流的在线健康社区（如好大夫、医享网等），以及呈爆炸式增长的医学科技文献等，见图 2。命名实体识别<sup>[13]</sup>、新词发现<sup>[14]</sup>等大数据技术使得从多源异构异质的健康医疗大数据自动智能发现临床医学实体成为可能；而卷积神经网络（Convolutional Neural Network, CNN）、双向长短时记忆网络（Bi-directional Long Short Term Memory, Bi-LSTM）、双向编码器表征（Bidirectional Encoder Representation from Transformers, BERT）等深度学习模型已成功应用于临床实体标准化过程<sup>[15]</sup>，实现自底而上的临床医学术语补充与更新完善，提升临床医学术语体系对健康医疗大数据的覆盖度与兼容性。

### 3.4 与国际临床医学术语体系映射与互操作

基于术语统一编码开展中文临床医学术语体系

与国际术语标准（如 ICD、SNOMED CT 等）的映射与互操作研究，促进中文医学术语体系与国际术语标准的融合共进，提高我国在全球医疗卫生信息标准化中的话语权与国际地位。

### 4 建设实践探索

经过 5 年多的建设实践探索，中国医学科学院医学信息研究所重点结合我国临床实践活动和电子病历类真实世界数据，参考 SNOMED CT、ICD 等国际临床医学术语体系，遴选《常用临床医学名词》、名词委医学名词、《国家基本药物目录》、国家标准《疾病分类与代码》（GB/T 14396 - 2016）等临床医学术语标准并将其作为素材的主要来源，通过计算机自动的术语同义归并与专业人员加工审核，初步建成具有自主知识产权和中国特色的中文临床医学术语体系。具体而言，术语体系覆盖疾病、解剖部位、临床表现、有机体、诊疗项目、技术和方

法、化学药品和生物制品、医用设备、器械和材料、临床医学物质、心理行为、环境地理、人口学及社会经济学特征、诊查对象共 14 个维度<sup>[16]</sup>，收录 20 余万个概念、60 余万个术语，以及临床表现、治疗药物、并发症等上百种细分语义关系类型。以肺炎为例，术语内容包括中英文术语、释义、上下位关系、细分语义关系、与 ICD - 11 映射等（非疾病类术语的微观结构与此一致，区别在于细分语义关系类型不同），见图 3。此外以呼吸疾病为例，完成了语义知识丰富与试点应用研究；在收集 1 000 万篇中文医学核心期刊数据、2 万份电子病历、8 000 条健康社区帖子的基础上，正在开展多来源健康医疗大数据的命名实体识别与融合深度学习的临床实体标准化，预期建成超过 100 万条中文临床医学术语的知识体系。未来将重点考虑从内容完善、中西医术语融合、更新机制、开放服务及应用推广等方面继续推进。

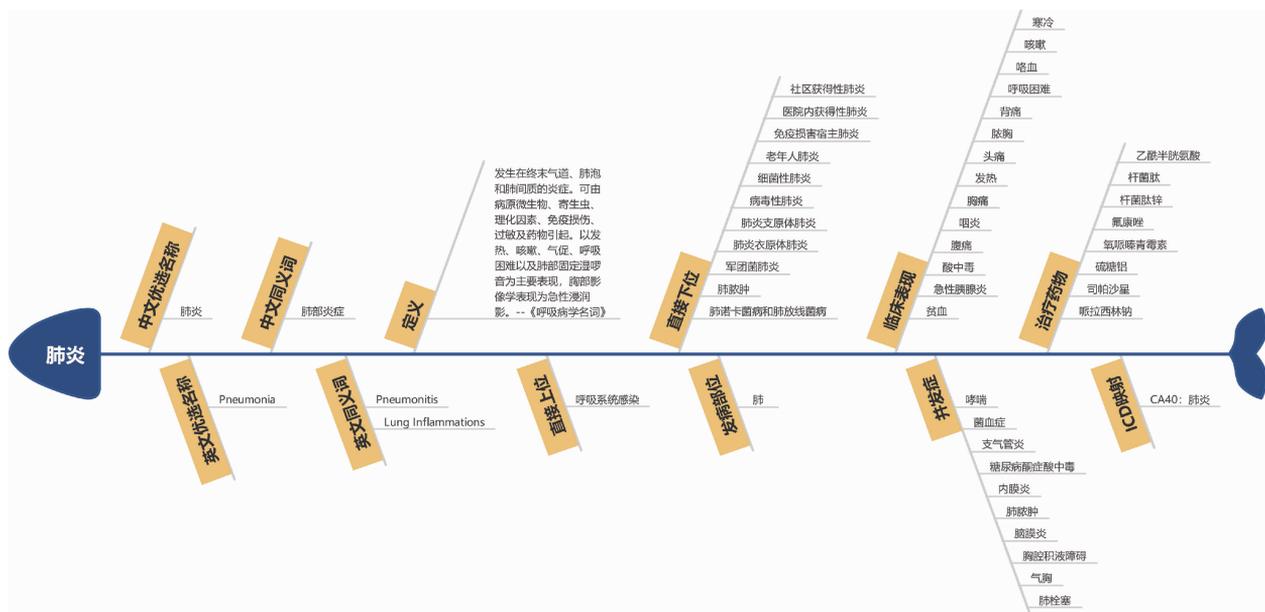


图 3 中文临床医学术语体系详细内容示例

### 5 结语

引进翻译与自主创新是建设中文临床医学术语体系两条不同的路径。中国医学科学院医学信息研

究所在多年医学术语标准理论与建设实践的基础上，初步探索出一条具有自主知识产权的中文临床医学术语体系建设之路，助力实现中文临床医学信息的规范表示、互通共享与挖掘利用，并对我国乃至其他母语非英语国家的医疗卫生信息标准化提

供借鉴。未来将在加强术语内容建设之余利用电子病历等现实世界数据进行质量验证,以及开展专科电子病历模板定制等应用研究。鉴于临床医学数据的复杂性与专业性,加之中文表达的多样性与多义性,在我国建设大规模、高度权威的临床医学术语体系依然任重道远。

## 参考文献

- 1 Chang E, Mostafa J. The Use of SNOMED CT, 2013 – 2020: A Literature Review [J]. Journal of the American Medical Informatics Association, 2021, 28 (9): 2017 – 2026.
- 2 Khorrami F, Ahmadi M, Sheikhtaheri A. Evaluation of SNOMED CT Content Coverage: A Systematic Literature Review [J]. Studies in Health Technology and Informatics, 2018 (248): 212 – 219.
- 3 Centers for Disease Control and Prevention. ICD HOME [EB/OL]. [2022 – 08 – 08]. <https://www.cdc.gov/nchs/icd/icd10.html>.
- 4 SNOMED International. SNOMED CT HOME [EB/OL]. [2022 – 08 – 08]. <https://www.snomed.org/>.
- 5 National Library of Medicine. RxNorm HOME [EB/OL]. [2022 – 08 – 08]. <http://www.nlm.nih.gov/research/umls/rxnorm/>.
- 6 Regenstrief Institute, Inc. LOINC HOME [EB/OL]. [2022 – 08 – 08]. <https://loinc.org/>.
- 7 ICH. MedDRA HOME [EB/OL]. [2022 – 08 – 08]. <https://www.meddra.org/>.
- 8 Bodenreider O, Cornet R, Vreeman D J. Recent Developments in Clinical Terminologies: SNOMEDCT, LOINC, and RxNorm [J]. Yearbook of Medical Informatics, 2018, 27 (1): 129 – 139.
- 9 任慧玲, 李晓瑛, 邓盼盼, 等. 国际医学术语体系进展及特色优势分析 [J]. 中国科技术语, 2021, 23 (3): 18 – 25.
- 10 李丹亚, 李军莲. 医学知识组织系统: 术语与编码 [M]. 北京: 科学出版社, 2019.
- 11 李振吉. 中医基本名词术语中英对照国际标准 [M]. 北京: 人民卫生出版社, 2007.
- 12 李晓瑛, 李丹亚, 钱庆, 等. 面向知识组织系统整合的英文同义关系自动发现算法研究 [J]. 现代图书情报技术, 2014 (5): 26 – 32.
- 13 王红, 王彩雨. 中文医疗命名实体识别方法研究综述 [J]. 山东师范大学学报 (自然科学版), 2021, 36 (2): 109 – 117.
- 14 李晓瑛, 蔡妙芝, 李军莲, 等. 面向文献组织的重大公共卫生事件知识图谱构建研究 [J]. 医学信息学杂志, 2021, 42 (3): 8 – 13, 19.
- 15 李明洋. 面向中文电子病历的临床术语标准化研究 [D]. 北京: 北京交通大学, 2021.
- 16 冀玉静, 孙海霞, 邓盼盼, 等. 中文临床医学术语系统分类框架设计原则与构建思路 [J]. 中华医学图书情报杂志, 2018, 27 (11): 27 – 32.

## 《医学信息学杂志》版权声明

(1) 作者所投稿件无“抄袭”“剽窃”“一稿两投或多投”等学术不端行为,对于署名无异议,不涉及保密与知识产权的侵权等问题,文责自负。对于因上述问题引起的一切法律纠纷,完全由全体署名作者负责,无需编辑部承担连带责任。(2) 来稿刊用后,该稿包括印刷出版和电子出版在内的出版权、复制权、发行权、汇编权、翻译权及信息网络传播权已经转让给《医学信息学杂志》编辑部。除以纸载体形式出版外,本刊有权以光盘、网络期刊等其他方式刊登文稿,本刊已加入万方数据“数字化期刊群”、重庆维普“中文科技期刊数据库”、清华同方“中国期刊全文数据库”、中邮阅读网。(3) 作者著作权使用费与本刊稿酬一次性给付,不再另行发放。作者如不同意文章入编,投稿时敬请说明。

《医学信息学杂志》编辑部