

医学术语标准化研究的思考^{*}

任慧玲 郭进京 孙海霞 李军莲 邓盼盼 冯 铭 王 怡

(中国医学科学院医学信息研究所
图书馆 北京 100005)

(中国医学科学院北京协和医院
北京 100730)

〔摘要〕 介绍国内外医学术语标准化研究现状, 分析医学术语系统建设流程和术语标准化研究中的关键问题, 采用系统分析方法对医学术语系统构建框架和流程进行探讨, 指出医学术语系统构建应着重解决医学术语遴选和素材库建设、概念体系建立和术语标准化、开放发布及应用维护问题。

〔关键词〕 术语标准化; 临床术语系统; 临床医学术语

〔中图分类号〕 R-056 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2018.05.001

Thinking of the Study on Medical Terminology Standardization REN Hui-ling, GUO Jin-jing, SUN Hai-xia, LI Jun-lian, DENG Pan-pan, Institute of Medical Information & Library, CAMS & PUMC, Beijing 100005, China; FENG Ming, WANG Yi, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences, Beijing 100730, China

〔Abstract〕 The paper introduces the current situation of study on medical terminology standardization in and outside of China, analyzes the building procedure of medical term system and key problems in the study on terminology standardization, discusses the building framework and procedure of medical terminology system with systemic analysis method, and points out that in the building of medical terminology system, several problems shall be focused on, including selection of medical terminology, building of material library, establishment of conceptual system, standardization of terminology, open release and application maintenance.

〔Keywords〕 Terminology standardization; Clinical terminology system; Clinical medical terminology

1 引言

术语 (Terminology)^[1] 是在特定学科领域用来表示概念的称谓集合。医学术语即是涵盖医学领域相关概念的称谓集合, 广泛分布在医疗记录、医学文献等信息资源中, 具有数量庞大、构词复杂等特点, 不同表达方式的缩略语、外来语以及大量存在

的一词多义、多词一义、含义不清 (歧义) 等现象阻碍健康医疗大数据的深度应用, 也影响医疗人工智能的效果。随着我国卫生信息化的深入, 对卫生信息标准要求更加迫切。虽然我国近年来加强了卫生信息化和标准化工作, 相继发布 280 多项卫生信息标准, 但是目前尚未建立规范、统一的医学术语标准。本文以中国医学科学院创新工程项目“中文临床医学术语系统构建研究”为基础, 探讨目前我国医学术语的构建框架、构建流程以及医学术语标准化研究中的关键问题。

〔修回日期〕 2018-05-21

〔作者简介〕 任慧玲, 研究馆员, 硕士。

〔基金项目〕 中国医学科学院医学与健康科技创新工程协同创新团队项目“中文临床医学术语系统构建研究” (项目编号: 2017-I2M-3-014)。

2 国内外研究现状

2.1 国外

国际上欧美等国医学信息标准发展较早,建立包括医学术语标准在内的标准体系。世界卫生组织(World Health Organization, WHO),欧洲标准化组织(European Committee for Standardization, CEN),美国国立医学图书馆(National Library of Medicine, NLM),国际医学术语标准化与研发组织(International Health Terminology Standards Development Organization, IHTSDO),医疗信息标准委员会(Health Care Information Standards Board, HISB)等积极从事医学术语相关标准的研究、制定、测试、批准、发布和推广活动^[2]。在各术语标准化组织的共同努力下,医学术语标准及术语集成果丰富,包括《医学术语系统命名—临床术语》(Systematized Nomenclature of Medicine—Clinical Terms, SNOMED CT)^[3],《医学主题词表》(Medical Subject Headings, MeSH)^[4],《观测指标标识符逻辑命名与编码系统》(Logical Observation Identifiers Names and Codes, LOINC)^[5],《国际疾病分类与代码》(International Classification of Diseases, ICD)^[6],《全球医疗设备命名》(Global Medical Device Nomenclature, GMDN)^[7]等,部分术语集完成了交叉映射和集成融合,如 SNOMED CT 完成与 ICD-10、ICD-9-M、LOINC 的交叉映射。这些术语体系在促进医学文献标引、检索、聚类与分析,各类卫生信息系统的卫生信息组织、描述、揭示与整合等方面发挥着重要作用^[8]。NLM 创建的一体化医学语言系统(Unified Medical Language System, UMLS)^[9]成为美国国家医疗卫生信息化的基础词汇系统。国际医学术语标准及术语集成果丰富,但由于存在版权、本地化、费用、语言表达等问题,我国的各类健康医疗数据用中文表达,具有中文语言和医学学科知识表达特点,国外术语系统不能直接满足国内临床信息化应用需求。

2.2 国内

近年来随着我国卫生信息化建设步伐加快,卫生信息标准体系的建立需要统一卫生领域各种术语信息标准和代码标准^[10]。同时健康医疗大数据已成为国家重要的基础性战略资源,为加强健康医疗大数据保障体系建设,应建立统一的疾病诊断编码、临床医学术语、检查检验规范、药品应用编码、信息数据接口和传输协议等相关标准,促进健康医疗大数据产品、服务流程标准化^[11]。医院内部结构化、智能化电子病历系统的建设,城乡居民规范化电子健康档案和健康卡的建立,对统一、规范化的医学术语有着迫切的需求。国内医学术语标准化建设起步较晚,许多国内的标准化组织和医学研究所、医学院校等相继参与到医学术语体系的建设中,包括中国标准化研究院、原卫生部医院管理研究所、中国医学科学院医学信息研究所、中国中医科学院中医药信息研究所、中国卫生信息学会及其下设的卫生信息标准委员会。这些组织和机构进行了大量国际标准的引进和应用,包括 ICD、LOINC 等术语及编码标准并开展标准的翻译、本地化研究和制定。目前国内医学术语标准体系的成果主要有《医学主题词表》(Chinese Medical Subject Headings, CMeSH)^[12]、《临床检验项目分类与代码》^[13]、《中国中医药学主题词表》^[14]、《中医药临床术语集》^[15]、《中医病症与分类代码》^[16]、《输血医学常用术语》^[17]等,这些成果在相关医学文献服务系统和卫生信息系统中得到应用。但是上述术语标准覆盖范围较为分散、体系不完善,缺乏一套完整的、涵盖临床诊断、检查、药物、疾病分类、症状描述等临床实践的中文临床医学术语标准体系。此外国内的医学术语标准体系建设还存在权威术语标准数量不足、质量参差不齐、网络化知识组织系统和工具贫乏、重制定轻实施等问题,不足以适应我国医疗卫生领域及行业的发展需求。因此建立面向临床应用、统一的中文临床医学术语系统及其共享服务体系具有重要的现实意义和长远价值,是推动医疗卫生信息化建设、计算机辅助临床决策支持以及人工智能在医疗领域深度应用的基础。

面,中文临床医学术语系统应覆盖疾病/诊断、发病部位、临床表现(症状和体征)、临床观察、检查、治疗、病理、化学品及药物、生物品、形态学等具有临床意义的不同医疗元素和相关元素之间的语义关联。在术语加工建设方面,搭建中文临床医学术语系统网络化工作平台,采用计算机辅助和专家审校相结合的方式,在概念归一、语义消歧、语义关系发现和一致性控制等计算机辅助构建关键技术研究的基础上,开展术语编辑、加工与审核。同时从互操作以及计算机管理的需求出发,制定中文临床医学术语编码标准,建立中文临床医学术语体系与其他术语体系(如 ICD-10)的映射,方便基于不同术语标准的异构卫生信息系统互操作。

3.2.3 发布层 面向我国电子病历、个人健康档案、国家医疗卫生服务、临床研究等应用需求,支持临床医学术语导航、检索、可视化、数据下载和第3方应用系统调用的开放共享服务平台和移动服务工具,提供中文临床医学术语系统开放共享。在开放共享的过程中确定术语使用方式和权益规范。

3.2.4 应用层 针对电子病历术语形式多样化和不规范现象,基于中文临床医学术语系统,可提供术语语义标注以及自动推荐等服务,实现电子病历的结构化、标准化存储,用于电子病历数据挖掘和临床决策支持。尤其是在当前大数据和人工智能快速发展的环境下标准化的临床医学术语可作为数据标准化和数据挖掘与推理的支撑,辅助临床诊疗决策。如针对国家新农合跨省就医联网结报数据多源、异构、标准化程度低的特点,本术语系统可提供针对疾病诊断名称、治疗方式、医疗设备等不同类型数据的术语语义标注及规范服务,实现跨机构、跨地区的疾病诊断相关分组规则提取、构建、分析和评估,辅助各级医疗卫生管理人员建立适宜本地医疗保险管理需求的疾病诊断相关分组器(Diagnosis Related Groups, DRGs)和管理工具。

3.3 关键问题

3.3.1 知识表示模型构建 按照术语体系建构原则,同时结合国内医学术语服务的需要,建立中文临床医学术语知识表示模型。建立在现代西医学本

体论疾病观基础上的 SNOMED CT 具有 19 类高层概念,涵盖身体结构、临床发现、观察对象、有机体、药物/生物产品、操作等。中医临床术语系统分类结构(Traditional Chinese Medicine——Categories of TCM Clinical Terminological System)将中医临床术语的顶层概念分为 17 个大类,涵盖症状体征、四诊对象、病症、中医操作/方法、病因病机等^[18]。中文临床医学术语知识表示模型的构建应当依据现代医学疾病知识体系,借鉴现有的以疾病为核心的知识组织体系(如 SNOMED CT、ICD-10、MeSH、《中图法》R 大类等),结合临床医学专家、术语专家的智慧,建立中文临床医学术语顶级类目和关系框架,对临床医学概念、术语的中文表达形式、语义关系等取值进行规范,建立以概念为中心的临床医学知识表示模型。

3.3.2 标准化 中文临床医学术语系统的建设应当保证其权威性、规范化、全面性,面向临床实践需要,从术语来源、描述形式、语种、内容质量等方面制定中文临床医生术语素材采集、遴选标准。从当前国内词典编制和出版的现状来看,中文临床医学术语系统的素材遴选应着重考虑中华医学会及其分会、全国科学技术名词审定委员会(医学)发布的医学名词标准,以及权威出版机构(如科学出版社、人民卫生出版社)出版的临床医学类书籍、指南、词典、操作手册等。遵循标准化管理原则,确定唯一的中文规范名称。按照《术语工作:原则与方法》(GB/T10112-1999)以及术语标准化的要求,术语定名时应该遵循如下原则:(1)专业性。(2)约定俗成性。合乎语言习惯,不引起歧义。(3)单义性。术语与概念一一对应,可以多个术语表达同一个概念,但是一个术语仅表达唯一的概念,不能有歧义。(4)科学性。注重术语的学术含义,准确表达概念的科学内涵和本质属性。(5)简明性。易懂、易记、易读、简洁。(6)稳定性。一经定名不宜轻易改动。(7)系统性。特定领域的各个术语必须处在一个层次结构明确的系统中,通过概念表达知识结构。(8)确切性。术语确切反映概念的本质特征。(9)国际性。与国际术语概念接轨^[19]。术语及其所表达的概念之间存在的关系可以

概括为层级关系和相关关系。层级关系主要包括属种关系、整体与部分的关系；相关关系一般可以概括为空间（位置）关系、时间关系、因果关系、源流关系、发展关系等。这些关系在医学知识体系中普遍存在。如神经系统肿瘤包括中枢神经系统肿瘤和周围神经系统肿瘤，中枢神经系统肿瘤又可以分为脑肿瘤、脊髓肿瘤等，这样在概念上就组成了包含关系，即属种关系。整体与部分的关系广泛存在于人体解剖结构类术语中，如垂体是下丘脑的一部分等。当然，医学术语是随着医学学科的发展不断更新，在术语建设过程中应充分结合临床医学专家、术语专家的智慧，做好术语审核和维护更新。

3.3.3 网络化加工平台建设 中文临床医学术语系统构建是一项浩大的工程，术语素材仓储建设、知识表示模型建设、术语的编辑、加工与审核应充分利用人机结合的方式以及自然语言处理和文本挖掘技术，采用机器学习和专家审核互相迭代的人机协同构建模式，构建集内容编制、质量控制、计算机辅助等为一体的协同工作平台和工具。

4 成果及意义

4.1 概述

建成之后的中文临床医学术语系统将在内容上覆盖疾病/诊断、发病部位、临床表现（症状和体征）、临床观察、检查、治疗、药物品、形态学等具有临床意义的不同医疗元素和相关元素之间的语义关联，能够较好地解决临床用语同义、多义及表达不规范问题，实现不同医疗元素之间的语义关联，推动我国医疗卫生信息化基础标准建设。中文临床医学术语系统主要面向科研、临床、教学人员提供服务，受益人群覆盖医务工作者、科研人员、社会公众、决策管理者等。

4.2 促进我国自主知识产权的临床医学术语标准体系发展

中文临床医学术语系统的构建充分结合我国临床实践过程中的术语标准化应用需求，是我国首个具有自主知识产权的临床医学术语实用体系，有助于改变

当前我国卫生信息标准和规范建设滞后的局面。

4.3 为我国医疗大数据智能技术研究和产业发展提供基础支撑

《全国医院信息化建设标准与规范（试行）》^[20]将人工智能、大数据、物联网和云计算等新兴技术摆在三甲医院信息化建设的最高位置，而人工智能、大数据在临床辅助决策、疾病分析与诊断方面的作用发挥需要实现数据的标准化、规范化。随着医疗信息化和电子病历的发展，医疗领域积累了大量以叙述性文本为主要形式的数据，需要利用医学语言处理技术处理和利用这些数据，以支持健康医疗大数据挖掘、临床决策和临床科技创新等。中文临床医学术语系统的构建为临床医学术语的标准化、规范化处理提供统一的术语标准体系，有助于实现病历文本的语义标识和结构化处理，为健康医疗数据挖掘提供知识组织工具。

4.4 服务国家人口健康战略，推动智慧医疗发展

《“十三五”规划纲要》明确指出要完善医疗服务体系，推动电子健康档案等工作开展^[21]。中文临床医学术语系统的建设能够实现医学术语的标准化，规范电子健康记录的书写及管理，使电子健康记录的术语更为精确，解决临床系统中电子病历、医嘱录入、监护病房、急诊室表格记录等之间的语言交互障碍^[22]，支持医院临床信息系统对患者诊疗信息的收集、存储、处理、提取等，实现不同来源、跨领域的信息共享、互联互通和医疗机构的协同服务。在利用医学术语标准体系规范化记录电子健康档案的基础上可以对健康数据进行统计分析，根据分析结果制定医疗卫生政策，推动智慧医疗发展。

5 结语

建立具有我国自主知识产权的、面向计算机应用、以概念为中心、能够揭示不同概念之间临床关系的中文临床医学术语系统具有重要的现实意义和长远价值，对电子病历规范化、健康医疗大数据的

应用和发展、满足我国临床信息化应用需求有积极意义,是支撑国家医疗卫生信息化建设的基础。中文临床医学术语系统构建应遵循规范、严谨的知识表示模型,借助人机协同加工平台,保证内容体系建设质量,以健全我国人口健康信息化和健康医疗大数据标准体系。

参考文献

- 1 Wikipedia. Terminology [EB/OL]. [2018 - 04 - 30]. <https://en.wikipedia.org/wiki/Terminology>.
- 2 钱庆,吴思竹. 国外医学术语标准化发展对我国的启示 [J]. 医学信息学杂志, 2013, 34 (5): 42 - 46.
- 3 SNOMED International. SNOMED CT [EB/OL]. [2018 - 04 - 30]. <https://www.snomed.org/snomed-ct>.
- 4 U. S. National Library of Medicine. Medical Subject Headings [EB/OL]. [2018 - 04 - 30]. <https://www.nlm.nih.gov/mesh/meshhome.html>.
- 5 LOINC. The Universal Standard for Identifying Health Measurements, Observations, and Documents [EB/OL]. [2018 - 04 - 30]. <https://loinc.org/>.
- 6 World Health Organization. International Classification of Diseases [EB/OL]. [2018 - 04 - 30]. <http://www.who.int/classifications/icd/en/>.
- 7 MedNet GmbH. The Global Medical Device Nomenclature [EB/OL]. [2018 - 04 - 30]. <http://www.mednet-eu-rep.com/gmdn.html>.
- 8 孙海霞,李丹亚,李军莲,等. 医学术语仓储管理系统构建 [J]. 医学信息学杂志, 2013, 34 (11): 48 - 54.
- 9 U. S. National Library of Medicine. Unified Medical Language System (UMLS) [EB/OL]. [2018 - 04 - 30]. <https://www.nlm.nih.gov/research/umls/>.
- 10 原卫生部 国家中医药管理局关于加强卫生信息化建设的指导意见卫办发〔2012〕38 号 [EB/OL]. [2018 - 04 - 30]. <http://www.moh.gov.cn/mohbgt/s6693/201206/55102.shtml>.
- 11 国务院办公厅. 国务院办公厅关于促进和规范健康医疗大数据应用发展的指导意见国办发〔2016〕47 号 [EB/OL]. [2018 - 04 - 30]. http://www.gov.cn/zhengce/content/2016-06/24/content_5085091.htm.
- 12 中国医学科学院医学信息研究所. 中文医学主题词表 [EB/OL]. [2018 - 04 - 30]. <http://cmesh.imicams.ac.cn/index.action?action=index&noMsg=1>.
- 13 原中华人民共和国国家卫生和计划生育委员会. 临床检验项目分类与代码 [EB/OL]. [2018 - 04 - 30]. <http://www.nhfpc.gov.cn/zwgkzt/s9492/201212/33995.shtml>.
- 14 中国中医科学院中医药信息研究所. 中国中医药学主题词表 [EB/OL]. [2018 - 04 - 30]. <http://tcmesh.org/>.
- 15 周霞继,刘保延,郭玉峰,等. 中医临床术语集在临床科研信息采集系统中的应用 [J]. 中国数字医学, 2010, 5 (10): 60 - 62.
- 16 国家中医药管理局. GB/T 15657 - 1995, 中医病证分类与代码 [M]. 北京: 中国标准出版社, 1995.
- 17 原中华人民共和国国家卫生和计划生育委员会. 输血医学常用术语 [EB/OL]. [2018 - 04 - 30]. <http://www.nhfpc.gov.cn/zwgkzt/s9493/201212/34407.shtml>.
- 18 董燕,李海燕,崔蒙,等. 中医临床术语系统建设概况与改进措施 [J]. 医学信息学杂志, 2014, 35 (8): 43 - 48.
- 19 冯志伟. 现代术语学引论 [M]. 北京: 商务印书馆, 2011.
- 20 国家卫生健康委员会办公厅. 关于印发全国医院信息化建设标准与规范(试行)的通知 国卫办规划发〔2018〕4 号 [EB/OL]. [2018 - 04 - 30]. <http://www.moh.gov.cn/guihuaxxs/s10741/201804/5711872560ad4866a8f500814dcd7ddd.shtml>.
- 21 第十二届全国人民代表大会. 中华人民共和国国民经济和社会发展第十三个五年规划纲要 [EB/OL]. [2018 - 04 - 30].
- 22 李莎莎,董燕,孟凡红,等. SNOMED CT 的应用现状及发展趋势 [J]. 中国数字医学, 2016, 11 (1): 100 - 102.