

基于本体的医院医学术语管理系统设计研究*

李莹莹 徐美兰

任冠华

查裕忠

(浙江数字医疗卫生技术研究院 (中国标准化研究院 北京 100191) (浙江数字医疗卫生技术研究院 杭州 311113))

王志勇

郑杰

(同济大学附属上海市第四人民医院 上海 200434) (浙江数字医疗卫生技术研究院 杭州 311113)

〔摘要〕 介绍基于本体的医学术语管理优势, 阐述基于本体的医院医学术语管理系统设计及应用场景, 分析不足之处并提出建议, 为提高医院信息系统语义互操作性、医疗数据质量管理与应用水平提供参考。

〔关键词〕 医学术语管理; 本体; 医院信息系统

〔中图分类号〕 R-058 [文献标识码] A [DOI] 10.3969/j.issn.1673-6036.2022.08.003

Design of an Ontology – based Hospital Medical Terminology Management System LI Yingying, XU Meilan, Institute of Medical – care Information Technology, Hangzhou 311113, China; REN Guanhua, China National Institute of Standardization, Beijing 100191, China; ZHA Yuzhong, Institute of Medical – care Information Technology, Hangzhou 311113, China; WANG Zhiyong, Shanghai Fourth People’s Hospital Affiliated to Tongji University, Shanghai 200434, China; ZHENG Jie, Institute of Medical – care Information Technology, Hangzhou 311113, China

〔Abstract〕 The paper introduces the advantages of ontology – based medical terminology management, expounds the design and application scenarios of the ontology – based hospital medical terminology management system, analyzes the shortcomings and puts forward suggestions, so as to provide references for improving the semantic interoperability of Hospital Information System (HIS) and medical data quality management and application level.

〔Keywords〕 management of medical terminology; ontology; Hospital Information System (HIS)

〔修回日期〕 2022-06-20

〔作者简介〕 李莹莹, 硕士, 战略咨询部副总监, 发表论文 8 篇; 通信作者: 郑杰, 常务副院长。

〔基金项目〕 中国工程科技发展战略湖北研究院 2019 年咨询研究项目“区域医疗大数据平台支撑健康湖北建设战略研究”(项目编号: HB2019C17)。

1 引言

面对日益复杂的医院信息系统, 医院健康医疗数据管理难度急剧上升。医学术语作为临床诊疗数据规范化管理“四统一”的重要一环, 既是实现语义互操作的基础, 也是促进健康医疗大数据挖掘与

智能化应用的底层支撑^[1]。但现有医学术语管理模式仍以单系统独立维护与使用为主,缺乏整体统筹。导致各系统采用的医学术语、字典和编码体系存在同一术语表达不一致、同一术语内涵不同、同一术语集版本迭代不同步等问题^[2]。关于医院术语管理系统建设模式主要有离散管理和统一管理两类^[3-5]。近年来基于本体的医学术语体系备受关注,例如最新的国际疾病分类与编码第11次修订本(International Classification of Disease V11, ICD-11)采用了本体思想^[6]。现有医学术语服务平台可支持多类术语的查询、浏览与映射管理,如3M HDD^[7]、Ontoserver^[8]、SnoChillies^[9],也有围绕特定术语体系的服务平台,如Snowstorm(SNOMED CT)^[10]、LOINC FHIR术语服务器^[11]。但基于本体医学术语服务系统的研究^[12-14]缺乏面向医院实际场景的设计。本文提出一种基于本体的医院医学术语管理系统设计,探讨该系统在医院医学术语管理中的应用价值,以期为提高医院信息系统语义互操作性、医疗数据质量管理与应用水平提供参考。

2 医院医学术语管理模式分析

2.1 概述

术语管理是指支持在特定主题领域创建和管理术语的组织单元、流程和工具的总和,其目的是确保在整个组织内一致地使用正确的术语^[15],包括收集术语、记录术语、对术语进行准确性审查以及术语分发和传播4个过程^[16]。实现医学术语的语义一致性是医学术语管理的核心。

2.2 现有管理模式存在的问题

2.2.1 离散管理模式 目前主流的医学术语管理多为离散模式,医疗机构从业务需求出发,针对各业务系统搭建相应医学术语管理模块,如医院信息系统中设有诊断、手术字典以实现数据记录的规范化。离散模式下系统术语管理模块建设不受其他系统限制并可灵活调整,在信息化建设水平不统一的初期发挥了重要作用。但该模式缺乏整体统筹,各系统生成的数据无法有效互联互通^[3],导致后期集

成化发展过程中改造成本高、难度大。

2.2.2 统一管理模式 随着医院信息系统的集成化发展,统一的医学术语管理理念逐渐形成,对独立于各系统的术语进行统一组织管理,形成共享字典^[4-5],实现子程序随调随用,有效避免了离散管理中的语义异构问题。目前已有部分医疗信息化厂商基于此模式进行系统开发,但主要是将子系统字典进行简单集合,尚未形成有效的术语管理模型,在术语应用拓展性上存在不足。

2.3 基于本体的医学术语管理优势

2.3.1 有效解决语义异构问题 本体是共享概念模型明确的形式化规范说明^[17],其本质是通过“本体”实现概念以及概念间关系的唯一性描述,从信息根源上规避语义歧义产生。该特性契合术语管理中对语义一致性的核心要求。现实场景中医学用语具有多样化特点,医学学科疾病发展产生了大量医学新词,如果简单地采用字典的结构化录入往往无法适应临床实际工作需要,但随意输入又必然造成语义歧义。通过本体这一媒介将表达相同语义的多个医学术语标引至同一概念下,从而能够在用户端提供较为开放的输入方式,又能在信息层面有效消歧,是一种更柔性的管理模式。

2.3.2 实现知识共享、重用和发现 对术语字典的编码化管理可在形式上消除医学术语语义异构问题,但其管理维度是线性、一维的,仅实现概念的归一。而本体的核心是一种模型,除了明确概念还包括对概念相互关系的表达,是一种网络化的管理模式。这使得基于本体的管理在向下兼容词表、分类表特性的同时,还具备知识层面的拓展性。通过利用医学本体概念模型对医学公认知识的逻辑抽象,能够实现医学知识在系统中的共享和重用^[18],例如本体中子代疾病概念通过继承父代疾病概念的发生部位、形态学改变等用于逻辑定义概念的属性关系,在数据分析中不需要重复建立关系,从而显著提升医疗数据分析与处理效率。基于本体的医学术语管理还具备从已有知识推理更深层次、隐性知识的能力,能够有效支持复杂检索与逻辑推理,为机器学习医疗数据中的知识提供可解释性支撑。

3 基于本体的医院医学术语管理系统设计

3.1 架构

基于本体的医院医学术语管理系统主要由医学术语资源管理、医学本体管理、医学术语服务 3 个子系统构成，并通过应用程序接口（Application Programming Interface, API）将术语以服务化的形式赋能于医院信息系统数据管理的全流程。系统设计强调从全院层面对医学术语资源的统一管理调用，从源头实现语义层面的数据规范化，提升语义互操作水平。设计核心是通过医学本体管理模块串联术语资源管理与术语服务，利用本体丰富的关系网络，赋予计算机更深的医学语义理解能力，能够更好地处理结构化与非结构化健康医疗数据，从而解决统一管理模式下术语应用的拓展性问题，见图 1。

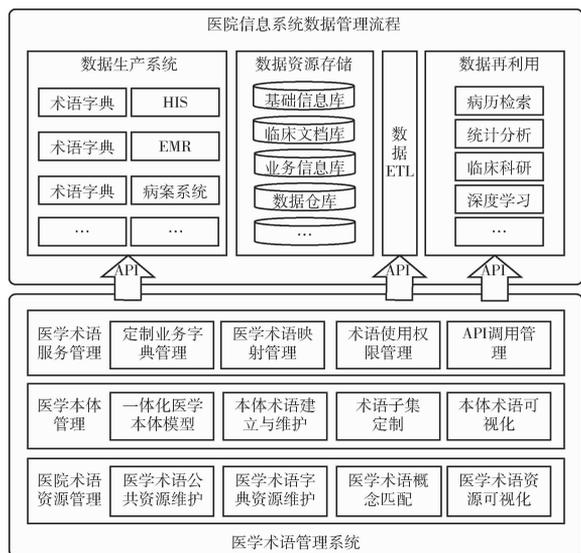


图 1 基于本体的医院医学术语管理系统架构

3.2 功能模块

3.2.1 医学术语资源管理 医院术语资源管理模块为医院提供多源异构术语资源的统一管理平台及各类术语向医学本体融合的工具集。通过收录疾病诊断编码、手术操作编码、药品分类及目录等编码标准规范和卫生信息标准中的值域类术语规范等各类医学术语标准、外部术语集，形成公共术语资源

池，满足医院对公共术语资源的查询、浏览、流转、调用以及版本维护更新需求。通过概念匹配将医院各类医学术语集映射到本体医学术语模型中，实现术语的统一管理。概念匹配是影响整个术语资源整合效率的关键，通过建立协作审核机制，并支持引入智能文本解析技术进行优化，见图 2。

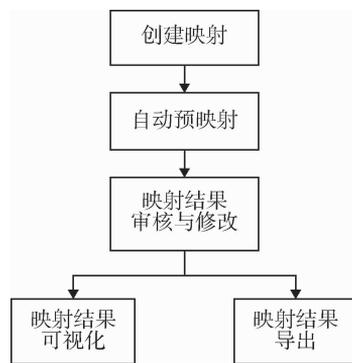


图 2 术语映射维护业务流程

3.2.2 医学本体管理 医学本体管理模块核心是建立医院一体化医学本体模型，为全院术语管理提供统一语义模型作为顶层框架。一般可选用现有相对成熟的医学本体模型为基础^[19]，通过本体模型的维护实现本体术语导入、本体术语中概念和概念的层级关系调整、概念的属性定义、概念中术语的增删改查、概念归一等功能，逐步完善医院医学术语本体模型。子集定制及拉取功能可从庞大的医学本体模型中抽取目标术语集，用于场景化数据挖掘。

3.2.3 医学术语服务管理 系统可在定制业务管理模块构建与维护定制业务字典。系统为每个业务字典配置调用 API，通过 API 调用实现不同系统间的字典同步与管理。本体映射维护功能支持结构化的术语数据实体间的映射，可实现不同术语的对齐或者合并，包括映射项目的创建与维护、智能预映射及映射结果审核、修改、导出等。通过权限管理针对不同业务单元、角色进行分级分类管理和功能开放设置，保障术语资源安全性与应用有序性。

4 讨论

4.1 系统应用场景探讨

4.1.1 ICD 编码管理规范化 ICD 编码是医院质

量管理、医疗保险费用结算等的重要编码体系。系统将 ICD 编码管理从原有的单个系统中抽离出来,从概念层面对编码进行统一维护、更新与版本映射转换,并建立 ICD 编码概念与临床医学术语概念的映射关系,各系统通过 API 调用形式实现对标准版本的使用与更新,从而满足全院层面的编码规范化要求。以手术编码为例,系统将全国各地 18 个版本的 16.2 万手术编码进行了语义层面的概念归一,形成 2.6 万概念量和 4.8 万 ICD-9-CM-3 编码术语集,医院可在此基础上建立内部统一的手术编码概念集及其与医学本体概念的映射,实现临床应用与编码的管理分离。在满足临床适用的诊断术语需求的同时能够通过映射关系实现多版本间的编码转换。

4.1.2 医学语义检索 构建病历搜索引擎时,针对不同医学文本特点可在术语管理系统中生成相应医学术语标注字典集,数据处理工具通过 API 调用标注字典集对临床文本进行更精细化的实体标注,例如利用本体模型识别病历文本数据中的“慢阻肺”是一种疾病、“阿司匹林”是一种药品、“艾滋病”“获得性免疫缺陷综合征”所表达的医学含义相同等。患者病历检索系统通过调用系统的医学本体一体化模型,可解析识别标注后的结构化临床病历数据中的属性、关系等,实现基于术语子集、术语映射、术语关联关系支持多条件的患者检索和相似病历查找等。例如查找“急性心肌梗塞”患者病历时,借助本体的同义词关系、层级关系,将具有“急性心梗”“急性心肌梗死”“急性前壁心梗”“急性后壁心梗”“急性多壁心梗”“急性广泛性心梗”等同样语义的患者信息都检索出来。

4.1.3 精确统计与医学推理分析 基于本体术语管理系统形成的医疗数据进入医疗数据分析系统时,可借助本体属性关系实现精确统计,例如利用病理分型属性关系精确统计诊断为“非小细胞肺癌”、病理分型为“鳞癌”,且使用了“吉西他滨”药物的患者数量及占到所有“非小细胞肺癌”患者的比例。本体医学术语形式化后的数据能够利用本体中蕴含的知识对医疗数据进行逻辑推理,包括层

级推理、属性推理等,提出可辅助临床诊疗或临床管理的决策,如根据患者症状推理出可能的疾病,帮助医生提高诊疗效率和质量。

4.2 系统不足之处与解决建议

4.2.1 术语管理信息模型构建难度大 术语管理既要有模型上的相对稳定性又需要有内容维护更新上的灵活性。这就要求在系统分析阶段准确地识别术语与系统、业务之间的相互关系,尽可能构建出适用于多源、多形态的术语管理以及应用可拓展的信息模型,确保搭建的系统架构能在较长时间内满足应用场景拓展需求。建议采用敏捷开发模式,通过快速迭代不断优化。在注重灵敏性的同时保证技术的标准化规范。

4.2.2 高并发应用下对 API 稳定性要求高 当真真正将术语及编码服务作为医院信息化基础支撑时,全院实时术语及编码服务 API 请求可能将达到每秒万次,因此对 API 并发性能的考虑也是建设的要点和难点。在建设需要结合实际规模,通过优化接口设计、请求分级队列控制、接口性能调整、服务缓存设计等方式优化以保障 API 服务的稳定性。

4.3 医院医学术语管理建议

4.3.1 重视术语管理体系建设 有效发挥医学术语管理系统效用需要管理层面的支撑。医院在信息化建设过程中,应建立全局性医学术语管理观念,将术语管理纳入信息化建设评估体系中。建立医学术语管理体系,逐步推动实现医院数据语义层面的规范化。从数据生产源头解决语义异构问题,提升数据质量,形成医院内医学术语规范化管理的“内循环”;形成统一的内外数据、跨机构数据交互模式,实现医院之间医学术语规范化管理的“外循环”。

4.3.2 多机构协同建立医学术语持续完善机制 医学本体建设依托于医学知识基础,随着现代医学知识的迭代更新,各专科领域的专业性更加突显,需要高水平团队协同构建,以确保术语模型的知识准确性。医学术语管理应形成多机构协调的可持续完善机制,减少重复投入,提高建设效率。

5 结语

基于本体的医院医学术语管理系统能够借助本体有效解决语义异构问题, 实现知识共享、重用和发现, 以及医院术语的统一管理与调用、编码管理规范化, 促进健康医疗大数据应用发挥价值。系统分析中应采用敏捷开发模式以逐步形成稳定适用的医学术语信息模型, 并通过优化 AIP 接口有效减轻高并发调用需求下对接口的调用压力。医院在推进医学术语管理系统建设的同时也应完善医学术语维护与管理机制建设。

参考文献

- 1 刘也良. 医疗服务正从“信息化”转向“智慧化” [J]. 中国卫生, 2019 (4): 84.
- 2 任慧玲, 郭进京, 孙海霞, 等. 医学术语标准化研究的思考 [J]. 医学信息学杂志, 2018, 39 (5): 2-7.
- 3 谭吉, 余立新. 当前医院信息化建设策略建议 [J]. 中国卫生信息管理杂志, 2018, 15 (6): 667-669.
- 4 何培根. 基于集成平台的医院基础数据字典建设 [J]. 医学信息, 2015, 28 (22): 2-3.
- 5 乐颖. 临床数据中心的构建与应用 [J]. 中国卫生产业, 2018, 15 (5): 149-150.
- 6 WHO. International Statistical Classification of Diseases and Related Health Problems (ICD) [EB/OL]. [2022-06-10]. <https://www.who.int/classifications/classification-of-diseases>.
- 7 3M. 3M Healthcare Data Dictionary: Controlled Medical Vocabulary Server [EB/OL]. [2022-06-10]. <https://multimedia.3m.com/mws/media/1759760/3m-hdd-overview-fact-sheet.pdf>.
- 8 Australian e-Health Research Centre, CSIRO. Ontoserver

- Terminology Server [EB/OL]. [2022-06-10]. <https://ontoserver.csiro.au>.
- 9 Valentia Technologies. Valentia Technologies Announces Launch of SnoChillies [EB/OL]. [2022-06-10]. <https://www.valentiatech.com/snochillies-launch.html>.
- 10 SNOMED International. Snowstorm Terminology Server [EB/OL]. [2022-06-10]. <https://github.com/IHTSDO/snowstorm>.
- 11 LOINC. LOINC Terminology Service Using FHIR [EB/OL]. [2022-06-10]. <https://loinc.org/fhir/>.
- 12 于琦, 李敬华, 李宗友, 等. 基于本体的中医医案知识服务与共享系统构建研究 [J]. 中国数字医学, 2017, 12 (5): 103-105.
- 13 杨虎, 王立军, 黄永峰. 基于本体的医学术语服务系统的设计与实现 [J]. 计算机工程与科学, 2013, 35 (12): 161-166.
- 14 王一杨, 王立军, 蒋嵩, 等. 面向医疗术语的本体库模型及其服务系统的设计 [J]. 计算机工程与设计, 2015, 36 (1): 258-262.
- 15 Kremer S, Kolbe L, Brenner W. Do You Know Your Terms? - A Procedure Model for Terminology Management [C]. Naples: Proceedings of the 11th European Conference on Information Systems, 2003.
- 16 Warburton K. Terminology Starter Guide [EB/OL]. [2022-03-31]. http://www.terminorgs.net/downloads/TerminOrgs_StarterGuide_V1.pdf.
- 17 Studer R, Benjamins V R, Fensel D. Knowledge Engineering: Principles and Methods [J]. Data & Knowledge Engineering, 1998, 25 (1): 161-197.
- 18 谢琪. 基于本体方法构建中医药概念信息模型的方法学示范研究 [D]. 北京: 中国中医科学院, 2011.
- 19 OMAHA. HiTA 术语 [EB/OL]. [2022-03-31]. <http://term.omaha.org.cn/>.

欢迎订阅

欢迎赐稿