

DOI:10.13288/j.11-2166/r.2024.11.005

思路与方法

中医知识库质量评价现状分析及指标构建

赵赛赛^{1,2}, 范雅祺¹, 张云³, 王维广¹✉, 陈子杰¹, 刘晓峰¹, 翟双庆¹

1. 北京中医药大学中医学院, 北京市朝阳区北三环东路 11 号, 100029; 2. 北京中医药大学东直门医院; 3. 电子科技大学

[摘要] 从中医资料类知识库和辅助决策类知识库两个角度, 梳理目前中医知识库建设、质量评价现状及其存在的问题, 在借鉴其他学科知识库质量评价策略和维度的基础上, 探讨中医知识库质量评价指标, 初步形成质量评价框架, 为中医知识库评价体系的完善提供思路。评价指标中, 基础性评价维度包括数据来源、数据整理、数据应用; 资料类知识库的特异性评价维度为数据质量, 辅助决策类知识库的特异性评价维度为数据匹配。各指标中, 数据应用维度中的数据量指标根据库中有效数据条目计数评价, 其他指标根据“是”“否”或“未知”定性评价后计分, 基础评价得分与特异评价得分相加得到总分, 根据分数将知识库分级, 结果以等级+数字的形式呈现。

[关键词] 中医知识库; 资料类知识库; 辅助决策类知识库; 质量评价; 评价指标; 质量控制

知识库是人工智能技术和数据库结合后的产物, 具备知识管理、演绎和推理能力, 可实现知识间交互共享和协作, 促进知识发现, 是研发专家系统、决策支持系统的关键^[1]。随着科技的发展, 知识库在构建过程中逐步融入本体和规则, 使知识库的性能从单一的数据仓储转向数据的深度利用。中医知识库是承载中医知识的重要平台, 用户借助中医知识库这一关键工具, 不仅能够系统地汲取已明确记载的显性知识, 还能积极探寻中医隐性信息; 知识库若使用先进的算法则能够为用户提供疾病的初步诊断、治法与方药的参考, 可为中医科研和临床提供有力支撑。

知识库质量主要包括数据质量、管理质量和用户体验等。知识库质量评价思路常围绕有用性、代表性、简单性三种属性^[2]展开, 其评估体系广泛涵盖技术性能、管理体系、用户满意度三个层面^[3]。目前我国知识库质量控制机制的建设处于起步阶段^[4], 质量评审体系尚未健全, 缺少普适性和实用性, 存在评价主体不明确, 标准不完善和流程不清晰等问题。中医知识库古文量大, 部分古籍数

据质量参差不齐, 质量与应用之间的矛盾不断凸显。合理的中医知识库质量评价体系能够在一定程度上对知识库的建设起规范引导作用, 促使提高数据质量, 促进知识挖掘, 实现以评促建, 减少中医知识库推广和流通障碍, 加速中医知识库标准化、国际化进程。因此, 健全中医知识库质量评价体系是提高知识库质量和实用价值、促进中医文化传承和传播的必然要求。

根据构建技术和内容应用方向^[1]的不同, 我们认为中医知识库可分为资料类和辅助决策类两种类型, 资料类知识库主要为文献知识库, 辅助决策类知识库则大多为临床知识库。本文从实用性的角度出发, 关注数据资源和应用需求, 结合两类知识库的共性与特异性, 从多个维度探讨中医知识库质量评价思路与指标, 为知识库研究人员提供参考, 助力中医知识库的共享和创新。

1 中医知识库建设与质量评价的现状存在的问题

1.1 中医知识库建设现状

中医知识库的建设始于 20 世纪 70 年代关幼波教授创建的肝炎诊治系统, 近年来知识库发展迅速, 现有种类繁多, 截至 2021 年, 在平台上集成的可检索数据资源已有 25 类, 部分已发布^[5]。中医知识库常基于本体、关系型数据库或神经网络算

基金项目: 科技部科技基础资源调查专项(2022FY102000, 2022FY102002); 中央高校基本科研业务费专项资金(2022-JYB-JBZR-011)

✉ 通讯作者: 202001031@bucm.edu.cn

法构建,技术手段包括计算机辅助切分词、知识抽取、语义分析、语义标注、知识图谱等。下面从资料类知识库和辅助决策类知识库两个方面分析中医知识库建设的现状。

1.1.1 资料类知识库 资料类知识库侧重于对领域中海量古今文献的管理和展现,其文本常为非结构化数据,具有知识储存、表示、分享等功能。例如,通过对古籍的整理和标注建立的首个中医古籍知识库系统^[6],涵盖了本草、方剂、医案等专题知识;选取46本具有代表性的温病古籍构建的知识库^[7],便于学者研究温病的学术特点和诊疗方法;将中医养生知识分类后构建的养生专题文献知识库^[8],使用户可以便捷地查询大量中医养生知识。资料类知识库体系以本体知识库、文献知识库与基础专题知识库为核心组成部分,其功能在于满足用户的各种资料查询需求。研发者在构建过程中首先设立框架,然后收集、抽取、融合信息,最后使信息规范地呈现、表达。该类知识库尚未融入诊断规则和推理机制,不具备辅助决策的性能,但部分可作为辅助决策类知识库的原始数据层。资料类知识库正处于迅速发展阶段,虽然在实际构建与运维过程中尚需面对一系列技术与学术难题,如术语统一标准的制定、知识内容的全面纵深覆盖以及各种概念间逻辑关联的界定等;然而,即便存在这些挑战,已有相当数量的资料类知识库成功投入运营^[5],并在整合信息资源、促进知识传播和推进人才培养等方面发挥着重要作用。

1.1.2 辅助决策类知识库 辅助决策类知识库是研发诊断辅助、用药辅助工具的基础,其文本多为半结构化数据或结构化数据,在经过知识抽取、标注、语义关联等数据搭建后,补充、丰富领域规则层、决策层或引入其他推理机制,提高知识库的演绎、推理能力,实现辅助决策的功能。例如,融入混合推理与解释机制的中医咳感症专家诊治样机系统^[9],其核心是涵盖领域内专家经验的知识库,是实现推理和解释的基础;将语义推理机制加入构建框架中的功能性胃肠病知识库^[10],既包括本体知识又实现了对疾病的早期诊断和药物推荐;应用贴近度算法的脾胃病本体知识库^[11],在定义诊断规则后延伸出一个小规模的脾胃病辅助诊断系统;作为中医智能处方推荐系统数据层的新型冠状病毒肺炎诊疗知识库^[12],在结合算法模型后能够自动推算合适处方,协助医师临床工作;其他知识库如冠心病知识库^[13]、中医骨伤知识库^[14]、哮喘病知识

库^[15]、中医医案知识库^[16]等均为此类。辅助决策类知识库汇集了大量中医经典著作、名老中医经验、疾病诊疗指南和临床实例等,旨在模拟中医临床思维过程,为临床决策提供智能辅助。此类知识库对人工智能相关技术要求较高,目前临床知识库大多处于实验阶段或理论建设阶段,在实现形成真正意义上的中医个体化、精准化诊疗建议方面仍存在一定困难。对辅助决策类知识库的不断探索和开发将有力推动公共卫生、健康管理、远程医疗服务的进步与发展。

1.2 中医知识库质量评价研究的现状与存在问题

中医知识库是中医药信息化建设的重要组成部分,属于数据知识库。有研究者^[3]梳理了我国数据知识库质量控制研究现状,提出我国目前相关研究滞后,处于初级阶段,数据质量和应用之间的矛盾日益尖锐,质量控制方法、流程模糊,数据质量评审体系不健全,极易出现低效率、主观性评价。亦有学者^[17]探索中医文献质量评价方法,采用中医古籍证据评价分级量表将古籍证据分为诊断和防治两类,依据证据的来源和内容来制定评价和量化条目。但较少学者关注知识库的数据质量和应用需求,关于中医知识库的质量控制、评价研究也较少。2023年5月21日,我们以“中医知识库”为主题检索中国知网,得到78篇相关文献,其中仅个别文献提及知识库的评价,且多是基于知识库构建技术^[1]、框架思路^[18]或某一组成要素^[17]的论述,未检索到关于知识库应用的综合质量评价的相关文献。

综上,中医知识库质量评价主要存在如下问题:1)缺乏统一、规范、普适、公认的评价标准;2)评价指标不能全面反映中医知识库的价值,缺乏对知识库实用性、科学性、创新性的判断指标;3)缺乏反映知识库更新、维护和优化能力的动态评价机制;4)评价标准多停留在理论层面,应用不足,实施效果和评价反馈缺失。

2 中医知识库质量评价思路与指标构建

2.1 知识库常用评价思路 and 维度

中医知识库质量评价体系的构建需借鉴其他领域知识库成熟的评价策略与模式,建立多元化视角深入评估知识库质量和性能的体系。知识库的评价思路普遍具有三个特性,即有用性、代表性和简单性^[2]。有用性指评估知识库是否满足实际需求,解决问题或填补信息空缺;代表性指评估知识库所涵

盖内容是否全面,是否平衡且可靠地代表相关领域知识;简单性指评估知识库是否易于理解和使用。如有学者^[2]从以上三个维度评估世界卫生组织全球血液安全知识库被开发为监测系统的实际潜力,对从实用性相关角度构思中医知识库评价指标具有启发意义。

知识库的质量评价框架通常涵盖技术、管理政策及用户三大维度^[3]。其中,技术层面确保数据质量,关注数据的准确无误、一致性和完整性等;管理政策层面聚焦于数据规模和版权合规等管理实践;用户层面则重视用户体验,如认知度和满意度。中医知识库质量评价体系也应具备综合维度,使评价体系的完善不仅是理论的深化,更具有以评促建的现实意义,即技术层面促进提升中医知识库底层架构,管理政策层面关注知识产权保护与内容质量把控,保证中医知识的权威性和可靠性,用户层面则通过改善用户感知、增加使用黏性,进而提升中医知识的社会认同度和使用普及率。

2.2 中医知识库质量评价原则和设计

在知识库常用的质量评价指标基础上,结合中医知识库建设和发展实际,从数据资源和知识库应用的角度初步探讨中医知识库质量评价思路与指标。

我们认为,建立中医知识库质量评价体系要坚持统筹兼顾、遵循务实准确的评价原则。1) 坚持普适原则。中医知识库涉及的子领域多,各知识库构建方法、技术不尽相同,为防止评价体系冗杂,将知识库分为资料类和辅助决策类,在共性评价的基础上,根据各自特点提出特异的评价指标。2) 坚持实用原则。区别于机构知识库,中医知识库不仅需要具备资料管理、检索性能,部分还需具有辅助决策潜力。资料类知识库重视文献资料的内容质量,对知识库构建技术的要求则相对宽松,而辅助决策类知识库借助先进的语义系统、学习模型才能够发挥更大价值,其对构建技术要求更高。质量评价体系的建立是为了加强知识库质量控制,因此,不同类的知识库评价维度、指标应各有侧重。3) 坚持实事求是原则。中医知识库涵盖大量文献,其版权有无争议在评价标准中应有所体现。

2.3 中医知识库质量评价维度

数据资源是知识库的核心,应用性能是影响用户选择的重要因素。因此,评价体系需体现数据资源的可靠性、准确性和实用性等。我们认为评价体系应分为基础性评价和特异性评价两部分,基础性评价适用于所有中医知识库,包括数据来源、数据

整理和数据应用三个维度。数据来源维度相关指标评判知识产权和知识库运行或依托主体;数据整理维度相关指标涉及差错率,体现知识库准确性;数据应用维度则涉及数据量和资源结构。特异性评价主要体现在数据质量和数据匹配上,资料类知识库侧重文献质量评价,辅助决策类知识库侧重匹配算法评价。

2.4 中医知识库质量评价的基础性指标

2.4.1 数据来源 知识产权贯穿知识库整个生命周期,关系着资源的收集和使用,影响知识库的推广和商业价值的实现。知识产权如果存在争议,不仅知识库内容权威性、可靠性受损,内容无法完全呈现,还可能会引起法律纠纷,降低知识库的使用质量。因此,设置该指标的意义是督促知识库建设者明确数据来源,关注数据版权问题。

知识库是否由权威机构、专业组织、学术团体等运行或是否被其认可、推荐,是衡量知识库可靠性的重要参考。例如,中医药古文文献知识库^[6]包括本草、方剂以及6个古籍专题知识库,2014年在中国中医科学院医史文献研究所支持下建立,并在多家中医药大学的帮助下联合完成,因而普遍认为中医药古文文献知识库信息源可靠、数据较准确。设置这一指标的目的在于考察知识库信息源是否可靠。获得权威机构推荐也便于知识库的营销和推广,可以提高用户对产品或服务的信任度和购买意愿。

2.4.2 数据整理 数据整理分为人工整理和机器整理两类。人工整理主要是借助随机抽样、人工校对、专家标注实现,机器整理则是利用数据管理和元数据创建等工具。例如,四字编码技术和借助光学字符(OCR)识别辅助人工校对技术是常用的整理方法,前者差错率在万分之一以内,而后者差错率在千分之一以内。合理的整理方式会提高数据集的质量和数据的准确率,使用户获得更好的使用体验,故设置“是否提供数据差错率结果”这一指标,清晰的数据差错率在一定程度上可以督促知识库创建者使用先进的数据整理技术,提高知识库数据准确率,对建设科学的知识库有重要意义。

2.4.3 数据应用 数据应用维度包括定量与定性两个指标,定量指标为数据量指标,直接反映知识库有效数据条目,定性指标则重点评判知识库文献资源内容是否符合主题、结构是否合理。

知识库存储的数据量是知识库价值的重要体现,存储量越多,应用者越能获得高质量的信息服务体验。例如,中医药古文文献知识库^[6]的数据资

源达 200 余万条, 其中方药知识库数据 20 余万条, 大数据量的知识库可以提供一站式检索, 能够给予用户更好的体验。此外, 对于各类临床知识库来说, 内容能否及时更新直接影响使用者是否能够查询到最新的疾病指南或诊治经验。数据量是一个可量化的动态对象, 因此可将数据量有效条目以数字的形式呈现, 直观表达知识库提供信息服务及其更新优化的能力。

知识库文献资源是否符合知识库主题、结构是否合理与知识库应用密切相关。不同中医子领域的知识库资源覆盖面不同, 侧重也不同, 知识库的知识结构合理, 内容切合主题, 用户在对知识库进行检索后可以获得想查询的知识, 满足使用需求, 实现知识库的应用价值。如用户通过检索温病古籍知识库^[7], 能够满足研究古代温病医家的需求, 符合知识库建设初衷; 辅助决策类的临床知识库, 知识资源需包括临床指南、名家经验和古今文献等, 应系统地涵盖临床治疗疾病所需要的参考内容。因此, 评判该指标需明确知识库主题与研究范围, 邀请专家对知识库文献资源抽样评审, 专家可以通过文献检索与验证、用户意见反馈、可视化的文献分布图等综合评定。设置该指标, 重视资源内容、结构, 关注用户应用需求, 旨在促进知识库应用价值的提高。

2.5 中医知识库质量评价的特异性指标

2.5.1 数据质量 资料质量是数据质量的重要组成部分, 主要考察构成知识库内容的基础文献、报告、文章等原始材料的质量, 是资料类知识库的竞争核心。文献质量评价是内容评价的一种。近年来, 学者不断提出和完善中医药证据体系和评价要素, 其衡量文献质量的思路可以借鉴用于知识库的文献质量评价。例如, 具有广泛认可度的中医古籍评分^[19], 其包括古籍引用量、作者及出处因素, 评分高意味着古籍的认可度高, 质量好。如果知识库中此类古籍占比多, 则知识库的数据质量更高。总之, 知识库的文献质量评价方法应为公开发行, 有一定认可度, 高质量文献占比应在文献质量评价中有所体现。本评价体系中设置“是否提供文献质量评价报告”这一个定性指标, 旨在增加知识库建设者对文献质量的重视程度, 提高知识库数据质量。

2.5.2 数据匹配 在知识库的构建中, 框架匹配或实体匹配^[20]至关重要, 影响着知识库的知识融合、获取、储存和推理。框架匹配包括元素匹配和结构级匹配, 通过算法能够判断出相似的概念、完

成元素之间的匹配, 解决数据的异构性问题。实体匹配^[20]也称实体对齐, 最初多基于语义相似性进行匹配, 近年来侧重实例匹配, 即判断两个实体是否指向同一对象。辅助决策类知识库重视知识库数据结构、规则层和知识推理。以功能性胃肠病知识库^[10]为例, 使用者输入“腹痛”“腹泻”“性情急躁”等关键词进行查询时, 系统能够精准关联到肠应激综合征这一诊断, 并为其提供详尽的中西医相关药物治疗方案参考。如果检索出的信息结构较完整、准确, 且没有信息冲突, 说明知识库中的匹配算法高效、知识融合较佳。因此, 针对辅助决策类知识库设置“是否提供实例匹配质量、效率相关报告”这一定性指标。相关报告应反映出知识库在文本数据方面的语义相似度水平、识别能力、容错机制与上下文信息感知能力等, 可以从查准率、召回率、缩减率等方面考虑, 前两者反映对齐的准确性和全面性, 后者反映候选匹配的筛选能力, 通过该指标可以引起业界对知识库匹配算法的重视, 满足用户精准检索需求。

2.6 中医知识库质量评价方法

中医知识库质量评价体系的评价主体为具有专业背景的第三方机构或学者。评价指标主要包括基础性评价和特异性评价两部分。基础性评价适用于所有知识库, 包括: 1) 数据来源; 2) 数据整理; 3) 数据应用。资料类知识库的特异性评价维度为数据质量, 辅助决策类知识库的特异评价维度为数据匹配。其中数据应用维度的数据量指标采取定量评价, 由数字构成, 数据库中每一个有效数据, 计 1 分, 上不封顶。其余指标根据表 1 进行评分, 采用“是”“否”或“未知”的定性评价, 结果为“是”计 2 分, 结果为“否”计 1 分, 结果为“未知”计 0 分。两部分得分相加即为该知识库的总分。总分 9~10 分为 A 级, 6~8 分为 B 级, 3~5 分为 C 级, 1~2 分为 D 级。知识库最终的评分为: 等级+数字的形式, 如有效数据计 11 000 个, 根据表格累计分值为 8 分, 8 分划分 B 等级, 则最终评分结果为 B (11 000)。

以中医症状近义词知识库^[21]为例, 其属于资料类知识库。根据表 1, 基础评价如下: 在数据来源上, 该知识库由北京中医药大学中医学院完成, 并于 2023 年 5 月获得计算机软件著作权认证, 因此其由专业团体运行, 知识产权无争议, 计 4 分; 在数据整理上, 该库尚未有公开的差错率结果, 计 1 分; 在数据应用上, 该数据库数据量上万条, 有效

表 1 中医知识库质量评价维度与定性指标

Table 1 Dimensions and qualitative indexes for quality evaluation of traditional Chinese medicine knowledge base

评价内容	评价维度	评价对象	指标
基础性评价	数据来源	资料类知识库和辅助决策类知识库	1) 知识产权是否存在争议 2) 是否由权威机构运行或由其认可、推荐
	数据整理	资料类知识库和辅助决策类知识库	是否提供数据差错率结果
	数据应用	资料类知识库和辅助决策类知识库	专家评审知识库文献资源是否全面包括符合知识库主题的文​​献内容
特异性评价	数据质量	资料类知识库	是否提供文献质量评价报告
	数据匹配	辅助决策类知识库	是否提供实例匹配质量、效率相关报告

数据条目为 17 240 个, 计 17 240 个, 其基本覆盖内科、外科、妇科、儿科等领域的所有医学症状词汇, 符合知识库主题且内容全面, 计 2 分; 特异评价如下: 该库尚未有公开的内容质量评价报告, 计 1 分。两部分得分总计 8 分, 根据该评分体系, 中医症状近义词知识库评分: B (17 240)。

3 结语

综上所述, 本文初步探讨中医知识库质量评价思路, 根据资料类和辅助决策类知识库的不同特点, 通过借鉴其他学科知识库的质量评价维度和指标, 设计通用维度包括: 数据来源、数据整理、数据应用, 特异维度包括: 针对资料类数据库的数据质量维度和针对辅助决策类数据库的数据匹配维度。各指标中, 数据应用维度的数据量指标根据库中有效数据条目计数评价, 其他指标采用定性评价后计分, 根据分数将知识库分级, 结果以等级+数字的形式呈现。评价结果可以作为用户选择知识库的参考, 加深用户对知识库的直观认识, 而完善知识库评价体系是知识库质量控制的重要步骤, 也是提高中医知识库竞争力的必然要求。

参考文献

[1] 牟梓君, 何丽云, 周雪忠, 等. 中医知识库的应用需求与构建方法分析[J]. 中国数字医学, 2021, 16(1): 35-39.

[2] KANAGASABA U, SELENIC D, CHEVALIER MS, et al. Evaluation of the WHO global database on blood safety [J]. Vox Sang, 2021, 116(2): 197-206.

[3] 张毅君. 机构知识库质量评价指标研究[J]. 现代情报, 2011, 31(10): 50-52, 56.

[4] 郑伯爽, 贺子岳, 陈晓峰. 数据知识库质量控制研究: 现状、问题与优化路径[J]. 中国传媒科技, 2022(8): 66-70.

[5] 王梦茹. 中医学专业知识服务系统资源发布[EB/OL]. (2021-12-10)[2023-10-02]. https://www.sohu.com/a/507037404_121124534.html.

[6] 杨继红. 中医古籍数字化资源建设概述[J]. 现代情报,

2008(5): 136-138.

[7] 李兵, 张华敏, 符永驰, 等. 基于语义关联的温病古籍知识检索系统的构建研究[J]. 辽宁中医杂志, 2012, 39(12): 2403-2404.

[8] 于琦, 于彤, 高宏杰, 等. 中医养生知识库构建方法研究[J]. 世界科学技术-中医药现代化, 2015, 17(8): 1612-1616.

[9] 陈国宁, 陈秋莲, 李陶深. 一个中医咳感症诊断专家系统的设计[J]. 广西大学学报(自然科学版), 2001, 26(2): 101-104.

[10] 朱玲, 李敬华, 于彤, 等. 基于本体的功能性胃肠病知识库构建研究[J]. 湖南中医药大学学报, 2018, 38(10): 1154-1160.

[11] 李新霞. 基于本体的中医学脾胃病知识库的构建[D]. 南京: 南京理工大学, 2008.

[12] 王斌, 刘涛, 王广志, 等. 支持新型冠状病毒肺炎的中医智能处方推荐和知识库系统[J]. 中国数字医学, 2020, 15(5): 25-27.

[13] 李硕. 基于 Jena 的中医冠心病本体检索的研究与实现[D]. 沈阳: 沈阳工业大学, 2016.

[14] 来建梅. 基于领域本体的中医骨伤知识库的研究与构建[D]. 济南: 山东中医药大学, 2013.

[15] 田野, 李敬华, 于彤, 等. 中医药防治哮喘病专题知识库构建[J]. 医学信息学杂志, 2017, 38(9): 65-67, 73.

[16] 翁晓兰, 申力, 贾佳, 等. 中医医案数据库的构建及其应用研究[J]. 浙江中医杂志, 2016, 51(8): 613-614.

[17] 张磊, 曾宪涛, 佟琳, 等. 运用循证医学原则制定中医古籍证据评价分级量表的思路[J]. 中华中医药杂志, 2020, 35(6): 2971-2973.

[18] 高曼, 李海燕. 中医药信息学应用研究热点[J]. 首都医科大学学报, 2022, 43(4): 592-599.

[19] 姚钰宁, 曹克刚, 托托, 等. 基于专家问卷的中医古籍证据分级及推荐方法的构建[J]. 中医杂志, 2021, 62(7): 572-576.

[20] 郭文龙. 中医方剂知识图谱构建研究与实现[D]. 兰州: 兰州大学, 2019.

[21] 北京中医药大学. 中医症状近义词数据库软件 V1.0: 2023SR0568649[P]. 2022-10-20.

(下转第 1125 页)