https://doi.org/bmr.202003.00051
此预印本（未经同行评审）的作者拥有该文稿的版权，biomedRxiv拥有永久保存权。任何人未经允许不得重复使用。

精准医学决策支持知识组织研究*

张秀梅[1]、翟运开[2]、丁楠[3]、吴晨溪[3]、张兵涛[3]、徐建武[1*]
（1 中国科学技术信息研究所 2 郑州大学管理工程学院 3 北京万方数据股份有限公司）

摘要：知识组织是临床决策支持系统（CDSS）的重要组成部分，相当于 CDSS 中枢神经系统。本研究通过综合分析国内外医学知识组织体系、深入研究 PM-CDSS 系统特点、知识概念及语义关系建设，构建适用于 PM-CDSS 的精准医学知识组织体系，切实提高临床决策系统诊断和治疗效果与效率。

关键词：知识组织，临床场景，精准知识库，临床决策支持系统，PM-CDSS

Abstract: Knowledge organization is an important part of clinical decision support system (CDSS), which is equivalent to the central nervous system of CDSS. Through comprehensive analysis of medical knowledge organization system at home and abroad, in-depth study of pm-cdss system characteristics, knowledge concept and semantic relationship construction, this study constructs a precise medical knowledge organization system suitable for pm-cdss, and effectively improves the diagnosis and treatment effect and efficiency of clinical decision-making system.

Keywords: Knowledge organization,clinical scenario,precise knowledge base, decision support system,PM-CDSS

1. 前言

临床决策支持系统（Clinic Decision Support System，CDSS）是下一代医院信息系统的重要组成部分，我国智慧医院及电子病历等相关政策也明确提出要求[1]。尤其随着互联网+、移动互联、大数据挖掘、人工智能、机器学习、区块链等新技术不断发展，医疗及医院信息化领域也面临着全新的挑战和机遇[2-3]。

CDSS 的历史可以追溯到 1976 年斯坦福大学研究开发的 MYCIN，近年来包括 IBM 沃森(Watson)引入到医疗领域，引起人工智能及医学界的关注和热议。虽然 CDSS 的历史发展久远，但现实应用中还存在诸如知识与规则流程融合问题、临床从循证医学走向精准医学问题、多学科联合决策支持问题、知识内容建设的严谨性和质量控质等问题[4-8]。

一个强大的 CDSS 离不开一个强大的知识组织支撑，本文从精准医学临床决策支持系统（PM-CDSS）出发，研究如何构建一个支撑 PM-CDSS 的精准医学知识组织体系，满足 CDSS 在多角色工作场景支持过程当中的信息需求，重点在于精准医学知识组织研究，关于 PM-CDSS 的系统设计将在另一篇文章中研究。

2. 知识组织相关研究

医学知识组织体系或知识组织系统全称（Medical Knowledge Organization Systems, MedKOS），涵盖医学领域内各种词表、概念体系、分类体系等，具有组织和描述医学知识、概念语义消歧和标准化、概念间层级结

*课题资助编号：国家重点研发计划“精准医学研究”重点专项（2017YFC0909900）基于远程/移动医疗网络的精准医疗综合服务示范体系建设与推广
*通讯作者：徐建武，中国科学技术信息研究所，112954897@qq.com

bmr.202003.00051V1

https://doi.org/bmr.202003.00051
此预印本（未经同行评审）的作者拥有该文稿的版权，biomedRxiv拥有永久保存权。任何人未经允许不得重复使用。

构和非层级结构语义关联展示等功能，属于图书情报学科领域中的一个分支，包括对知识单元或概念内在结构所进行的加工、整理、序化、表示等一系列操作。国外关于医学知识组织研究较为成熟，其中在临床及学术领域应用较为广泛的有：医学主题词表（MeSH）、一体化医学语言系统（UMLS）、系统化人&兽医学术语表临床术语（SNOMED CT）、观测指标标识符逻辑命名与编码系统（LONIC）、国际疾病分类（ICD）等；精准医学相关知识组织有：人类孟德尔遗传在线（OMIM）、罕见病（ORPHA）、中文人类表型（CHPO）、药物基因组学（PharmGKB_drugs）、基因（HGNC）、变异（DisGeNET）、药物编码查询系统（ATC）等；国内医学知识组织体系有：中国图书馆图书分类法、医学主题词表中文版、UMLS 中文版、中医药主题词表等[9-17]。

表 1 国内外主要医学知识组织体系

名称	介绍
医学主题词表（MeSH）	MeSH 是医学领域标准化叙词表，由美国国立医学图书馆 NLM 编制，也是动态可扩充的叙词表，汇集两万多个主题词。
一体化医学语言系统（UMLS）	UMLS 由美国国立医学图书馆编制，包括超级叙词表（Matathesaurus）、专家词典（Specialist Lexicon）和语义网（Semantic Web）
系统化人&兽医学术语表临床术语（SNOMED CT）	Systematized Nomenclature of Human and Veterinary Medicine Clinical Terms，即系统化人&兽医学术语表临床术语，是全面的临床医学相关术语集合，包含了词库、受控词表、分类系统、代码系统、语义网络，助于临床信息系统的检索和病历信息的语义互操作。
观测指标标识符逻辑命名与编码系统（LONIC）	LONIC 术语涉及用于临床医疗护理、结局管理和临床研究等目的的各种临床观测指标，如血红蛋白、血清钾、各种生命体征等。
国际医学分类族 WHO-FIC	①ICD(International Classification of Diseases)是世界卫生组织制定的国际国际疾病/健康分类,是依据疾病的某些特征，按照规则将疾病分门别类，并用编码的方法来表示的系统。 ②ICF(International Classification of Functioning, Disability and Health)是世界卫生组织制定的国际功能-障碍-健康分类，核心概念用于健康领域的良好状态，健康状况的限定语表现。 ③ICHI(International Classification of Health Interventions)，世界卫生组织制定的国际健康干预分类，提供一种用于统计目的并且对健康干预的分布和演变情况加以分析和报告的通用工具。
精准医学相关词表	①基因本体（Gene Ontology，GO）：是一个在生物信息学领域中广泛使用的本体，它涵盖生物学的三个方面：细胞组分、分子功能、生物过程。 ②在线人类孟德尔遗传数据库（OMIM）：涵盖关于人类遗传病和基因座位等相关信息和文献的中心级数据库。 ③罕见病（Orphanet）：关于罕见病诊断、治疗的专业网站，包含孤儿药和疾病信息。 ④人类表型术语集（HPO/CHPO）：旨在提供人类疾病中用于描述表型异常的标准词汇，其中文版为 CHPO。 ⑤遗传药理学与药物基因组学数据库（PharmGKB）：收集最完整的与药物基因组相关的基因型和表型信息，把基因和药物的关系分为两个大的范

bmr.202003.00051V1

https://doi.org/bmr.202003.00051
此预印本（未经同行评审）的作者拥有该文稿的版权，biomedRxiv拥有永久保存权。任何人未经允许不得重复使用。

	畴。 ⑥人类基因组数据库（HGNC）：对于每个 human 的基因来说，在 HGNC 数据库中会有一个数字标识的唯一 ID，也会有一个唯一的 symbol。 ⑦疾病相关的基因与突变位点数据库（DisGeNET）：一个专门收录人类疾病相关的基因与突变位点信息的数据库 ⑧药物编码查询系统（ATC）：即解剖学治疗学及化学分类系统，是世界卫生组织对药品的官方分类系统。
中国图书馆分类法	《中图分类法》是《中国图书馆分类法》的简称，是我国通用的图书分类工具。
中文医学主题词表（CMeSH）	CMeSH 由中国医学科学院医学信息研究所翻译、出版和维护,包括所有主题词、印刷款目词及副主题词的中译名
中文一体化医学语言系统（CUMLS）	CUMLS 由中国医学科学院医学信息研究所开发研制，由词表构建与维护平台、医学词表、语义网、词表服务平台组成。
中医药学主题词表	中医药学主题词表由中国中医科学院中医药信息研究所研制，是我国第一部中医药专业词表，在中医药学领域应用广泛。

3. 精准医学知识组织体系设计

3.1. PM-CDSS 系统概述及特点

基于场景的精准医学临床决策支持系统（Senario based clinical decision support system for precision medicine ,PM-CDSS）不仅满足精准诊断和精准治疗上提供决策支持，还包括场景的精准，对象的精准，时间的精准、决策信息的精准等，在情报学中可以概括为在恰当的时间、恰当的地点、将恰当的信息、以恰当的方式提供给恰当的需求者。因此构建一个覆盖全学科、多角色（支持医生、护士、药剂师、麻醉师、检查/验技师等）主要工作场景，满足诊前、诊中及诊后全流程的精准医学决策支持系统，为临床工作提供及时、有效全程的精准参考建议，进而降低医疗风险，提高总体诊疗水平是本研究目的[18]。本文则基于上述对于精准医学的理解和认知，重点分析 PM-CDSS 的场景需求及所对应的知识组织设计，有关决策支持的系统架构设计和场景应用将在另一篇论文中讨论。

PM-CDSS 特点体现在：对知识数据颗粒度要求更细、内容时效和准确度要求更高，需要抽提各类医学概念类；在推理过程中以医学概念类之间的语义关系为基础，建立语义规则库，为输出最佳决策信息提供支撑；并且能够适应不同终端载体，深入场景与各医院信息系统深度整合，动态精准的提供决策支持。

表 2 传统 CDSS 与精准 CDSS 区别

特点	传统 CDSS	精准 CDSS
数据来源	循证证据	循证精准知识
颗粒度	内容片段	碎片化知识
数据关系	无或弱关联	强关联规则
数据展示	静态阅读	动态提醒
数据调用	主动查阅	场景嵌入

3.2. 精准医学知识数据来源

精准医学知识数据来源一方面遵从循证医学要求，知识数据必须有权威可靠出处，才能够为决策支持所用，另一方面数据广度和深度要求更高，

bmr.202003.00051V1

https://doi.org/bmr.202003.00051
此预印本（未经同行评审）的作者拥有该文稿的版权，biomedRxiv拥有永久保存权。任何人未经允许不得重复使用。

数据广度方面增加了精准医学相关术语和知识，深度方面需要对数据进行二次碎片化加工，为临床决策做支撑。

表 3 精准医学主要概念及数据来源

知识库类型	内容	来源
疾病知识库/诊断知识库	临床问题	症状、体征、ICD10/11 诊断术语
	诊断推理规则	临床路径、相关指南中抽提
治疗知识库	检查方案	临床路径、临床指南、用药说明中提取
	治疗方案	
药物知识库	药品说明书	药品生产厂家、药典及其他途径
	合理用药	药物相互作用相关书籍及文献
	抗菌药物	国家和地方抗菌药物指南
检查检验及其他	检查检验知识库	临床检查检验操作指南等相关工具书
	正常值及异常值	
临床指南文献库	公开权威指南	卫健委网站及期刊杂志公开发表的正式指南
临床路径库	公开临床路径	卫健委网站
术语知识库	临床术语词典	SnomedCT、ICD10/11,ICF、ICHI、LOINC、UMLS、MeSH、CMeSH、OMIM、ORPHA、CHPO、PharmGKB、HGNC、DisGeNET 及非标准术语和值集等
外部数据	临床病例	基于临床病例数据挖掘的经验以及临床医生形成共识的经验
	临床经验	

精准医学知识数据来源可以通过以下方式获取：①国内外主要医学知识组织术语集合包括精准医学相关术语体系，如 ICD10/11、SNOMED CT、OMIM、ORPHA、CHPO 等；②国家颁布的各类标准，医保目录，医疗服务操作项目等；③临床路径或指南等正式公开出版发表的文献，权威工具书教材等；④官方正式渠道获取，如药品说明书等；⑤基于医院临床实际情况经过验证论证后的方法结论[19-20]。

此外，在以上数据采集的基础上，还需要对获取数据进行二次深度加工，包括数据碎片化加工，术语统一等处理，以及最终数据的审核入库。

3.3. 精准医学相关概念抽提

传统知识库以文献为单元进行知识组织，而精准医学知识库的数据颗粒度更小，需要将医学领域中涉及到的医学概念进行抽提标注。通过提取概念不仅可以为后续构建语义关系提供基础，不同的概念类别也能够为不同决策场景提供数据支持。

本文以临床医学指南、临床路径、检查检验操作指南、药物说明书等数据为基础，根据临床应用场景需求，将医学概念进行聚类归纳梳理后，总结出以下 10 个主要知识类型：①疾病类；②症状体征类；③病因病机类；④药物类；⑤检查类；⑥检验类；⑦护理类；⑧诊断类；⑨治疗类；⑩个体特征类。除定义每个概念类别自有属性值（如名称，别名，ID 等）外，还需定义概念相关属性，例如：数值属性，如年龄、性别都是单值属性，参考值范围是数值属性；对象属性指两个类别之间的

bmr.202003.00051V1

https://doi.org/bmr.202003.00051
此预印本（未经同行评审）的作者拥有该文稿的版权，biomedRxiv拥有永久保存权。任何人未经允许不得重复使用。

关系，如症状所发生的部位、适应症对应的疾病等[21-24]。

此外还需要对每个概念类赋予唯一 ID，建立统一的分类体系，并将其与院内各类医学术语词典进行匹配关联，便于后期语义关系的梳理，导航浏览及系统对接应用。

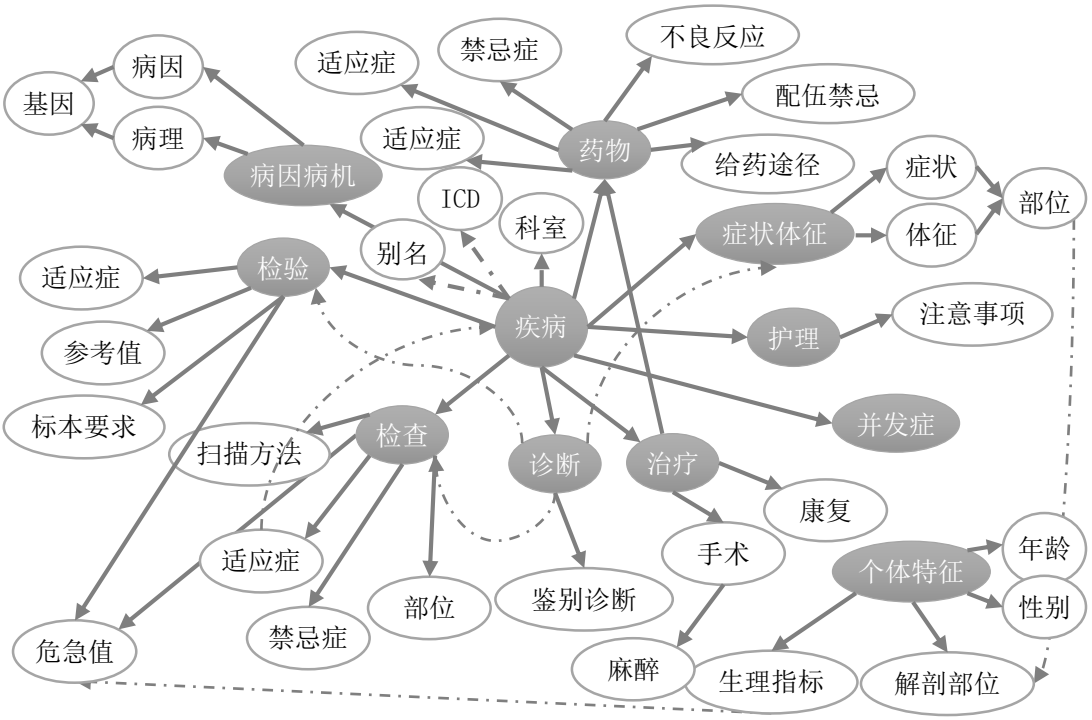

图 1 精准医学知识组织体系

表 4 精准医学知识语义关系

关系类型	示例	关系类型	示例
ID	ICD-10：C50.900	分期	ⅡA 期 分期 乳腺癌
类	乳腺癌 是 肿瘤科 乳腺癌 是 乳腺疾病	并发	放射性肺损伤 并发于 乳腺癌
名称	乳腺癌	参考值	癌抗原 15-3 测定 参考值 25～28U/ml
别名	乳腺癌 别名 breast cancer	危急值	血清肌酐 危急值 > 880μmol/L
致病	Her2 阳性 致病 乳腺癌	标本	癌抗原 15-3 测定 标本 血清
多发人群	乳腺癌 多发人群 女性	部位	乳腺 MRI 检查 部位 胸部
发生于	乳腺癌 发生于 胸部	服用方式	枸橼酸他莫昔芬 服用方式 口服
是症状	乳腺肿块 是症状 乳腺癌	服用间隔	枸橼酸他莫昔芬 服用间隔 1 日 1 次
诊断	乳腺细胞学检查 诊断 乳腺癌 乳腺 X 检查 诊断 乳腺癌	不良反应	枸橼酸他莫昔芬 不良反应月经周期紊乱
鉴别	乳腺良性增生病 鉴别 乳腺癌	禁忌症	枸橼酸他莫昔芬 禁忌症 妊娠期妇女
适应症	枸橼酸他莫昔芬 治疗 乳腺癌	配伍禁忌	阿莫西林 配伍禁忌 糖皮质激素

bmr.202003.00051V1

https://doi.org/bmr.202003.00051
此预印本（未经同行评审）的作者拥有该文稿的版权，biomedRxiv拥有永久保存权。任何人未经允许不得重复使用。

3.4. 精准医学知识语义关系

语义关系的构建是精准医学决策支持的关键，智能化的动态提醒离不开强大的数据规则库作为支撑，不同场景下通过调用不同的规则语义关系，可以实现不同的支持目的，如医嘱推荐、医嘱合理性判断、诊断及鉴别诊断等。

根据已有知识组织体系的相关研究[25-26]，归纳总结了如下常见的语义关联关系：1）等级关系：是一种、是成员、是实例。2）非等级关系：①等同关系：同义关系、近义关系；②相关关系：物理相关于、空间相关于、时间相关于、概念相关于、功能项关于（影响、并发、导致、治疗……）③其他。

本研究在前面已梳理的医学概念类的基础上，通过总结参考临床医学指南、临床路径、检查检验操作指南、药物说明书等知识中关于概念与概念之间的描述，梳理标注了 22 种精准医学知识语义关系，此 22 种语义关系为现阶段归纳总结的语义关系，由于医学知识组织体系是一个不断成长的体系，医学概念及语义关系也会不断更新调整。

4. 精准医学知识组织应用

关于知识组织体系的应用，其范畴较广，在信息检索优化、知识库构建、临床决策支持、命名实体识别、数据挖掘等方面都可以发挥巨大的作用。本文仅在场景化的精准医学决策支持方面进行讨论：

临床诊疗活动是一个连续的过程，单场景应用包括临床辅助诊断、临床诊疗建议、合理性检查、数值提醒、数据挖掘等方面，但是单场景应用方式无法完全发挥医学知识组织在临床实践中的应用效果，如仅有诊断结论而没有后续的治疗建议，或者仅有标准化的治疗方案但缺少相应的个性化用药指导等。在此基础上，提出基于临床业务流程的决策场景，在临床业务流程模型中，每个业务节点都与知识库进行关联，在不同节点进行相应的提醒，以门诊和住院两个场景分析为例：

4.1. 门诊场景

①询问病情时能够提供问诊及分诊信息；
②初步诊断时能够提供辅助诊断建议；
③下达检查检验医嘱时，能够推荐合理的检查检验并对其合理性进行判断；
④阅读结果报告时能有相关解读及危急值提醒标注；
⑤下达用药处方能够提供药物说明并合理性判断；
⑥患者离开医院后能够进行健康管理。

4.2. 住院场景

①撰写电子病历时有相关提醒及校正；
②初步诊断时能够进行鉴别诊断提醒；
③下达医嘱时有推荐医嘱提醒，并且对医嘱内容进行合理性判断及风险提醒；④阅读结果报告时能有相关解读及危急值提醒标注；
⑤制定并执行治疗处理方案时有风险评估、护理方案及相关风险预测提醒；
⑥患者出院后能够进行健康管理。

bmr.202003.00051V1

https://doi.org/bmr.202003.00051
此预印本（未经同行评审）的作者拥有该文稿的版权，biomedRxiv拥有永久保存权。任何人未经允许不得重复使用。

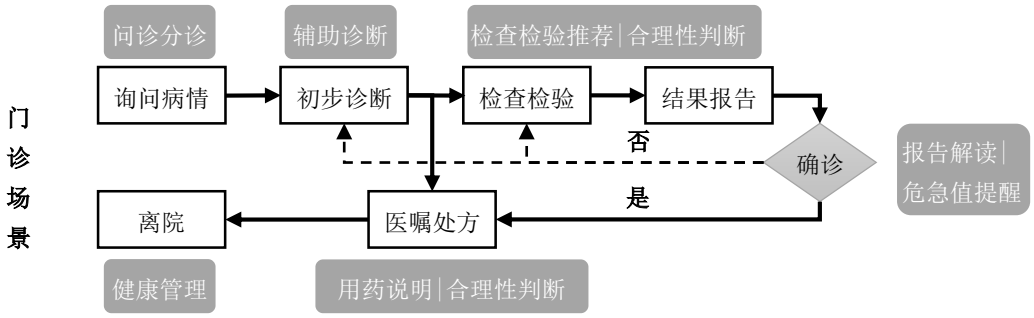

图 2 门诊场景精准医学决策支持

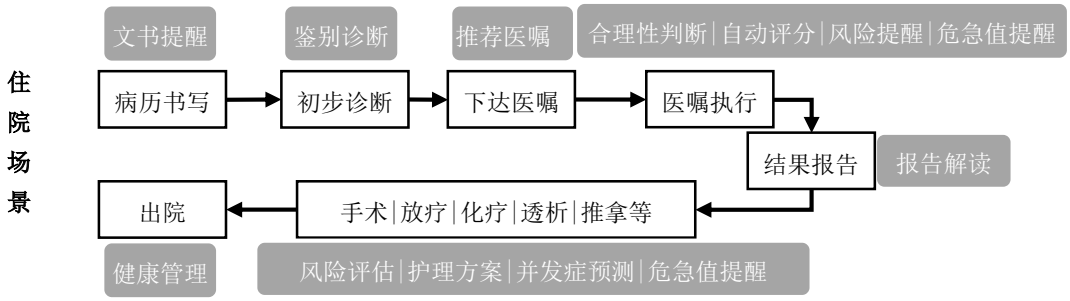

图 3 住院场景精准医学决策支持

5. 结语

基于场景的精准医学知识体系的研发，期望以临床业务流程为中心，满足在临床决策支持当中各种场景的需求。本文通过调研国内外医学知识组织体系及临床主要业务场景，对 PM-CDSS 的系统特点、医学知识组织体系、知识数据来源、概念抽提、语义关系梳理及场景应用进行了系统性研究，由于精准医学知识组织体系是一个复杂且庞大的工程，现有的知识组织体系距离在成熟 PM-CDSS 中应用还存在很大差距，下一步我们将在 PM-CDSS 的系统实现、知识组织的工程化以及知识组织体系的进一步完善方面继续开展研究。

bmr.202003.00051V1

参考文献

[1]http://www.nhc.gov.cn/yzygj/s7659/201812/3cae6834a65d48e9bfd783f3c7d54745.shtml
[2]李国磊,陈先来,夏冬.面向临床决策的电子病历系统概述[J].中国数字医学,2014,9(12): 30-32.
[3]王昱.基于电子病历的临床决策支持研究[D].杭州:浙江大学,2016.
[4] Sim I,Gorman P,Greenes RA,etc.Clinical decision support systems for the practice of evidence-based medicine.[J].2001,6(6).
[5]周娜,李爱芹,刘广伟,张国庆,张晓春.沃森肿瘤人工智能系统在临床中的应用[J].中国数字医学,2018,13(10):23-25.
[6]洪建,张洁,陶敏, 等.新型临床决策支持系统在智慧医院建设中的应用研究[J].中国数字医学,2019,14(9):22-24. DOI:10.3969/j.issn.1673-7571.2019.09.008.
[7]王宇,王心慰,刘爽, 等.精准医学的临床部署:顶层架构设计及关键信息技术[J].转化医学杂志,2017,6(6):321-324. DOI:10.3969/j.issn.2095-3097.2017.06.001.
[8]杨春华,王天津,黄思敏, 等.支持精准医疗的国外临床决策支持系统[J].中华医学图书情报杂志,2016,(2):14-19. DOI:10.3969/j.issn.1671-3982.2016.02.004.
[9]杜也力. 我国关于知识组织的研究述评[J]. 中国图书馆学报, 2002, 28(5):66-69.
[10]蒋永福, 李景正.论知识组织方法[J].

https://doi.org/bmr.202003.00051
此预印本（未经同行评审）的作者拥有该文稿的版权，biomedRxiv拥有永久保存权。任何人未经允许不得重复使用。

中国图书馆学报, 2001, 27(1):3-7.
[11]李洁, 丁颖. 语义网、语义网格和语义网络[J]. 计算机与现代化, 2007(7):38-41.
[12]牟冬梅, 崔艳玲. MeSH、本体论在医学知识组织中的作用[J]. 情报杂志, 2005,24(7):120-122.
[13]Cimino J J. Auditing the Unified Medical Language System with Semantic Methods[J].Journal of theAmerican Medical Informatics Association Jamia, 1998, 5(1):41.
[14]Stearns M Q, Price C, Spackman K A, et al. SNOMED clinical terms: overview of thedevelopment process and project status[J]. Proceedings / AMIA. Annual Symposium. AMIASymposium, 2001, 8(1):662.
[15]An Introduction to the Gene Ontology. The Gene Ontology[EB/OL].[2010- 12-01].http://www.geneontology.org/GO.doc.shtml.
[16]张晓梅, 李丹亚, 胡铁军. 一体化医学语言系统与本体论研究[J]. 医学信息学杂志,2006, 27(2):89-92.
[17]Mcdonald C J , Huff S M , Suico J G , et al. LOINC, a universal standard for identifying laboratory observations: a 5-year update[J]. Clinical Chemistry, 2003, 49(4):624.
[19]姜军,唐鹏.乳腺癌精准诊疗临床决策支持系统的关键技术[J].中华医学杂志 ,2015,95(31):2503-2505. DOI:10.3760/cma.j.issn.0376-2491.2015.31.003.
[19]张秀梅,徐建武,程煜华,等.基于知识库的临床决策支持系统构建[J].中华医院管理杂志,2014,30(6):472-475. DOI:10.3760/cma.j.issn.1000-6672.2014.06.025.
[20]李艳,吴梦佳,张士靖.语义互操作标准在临床决策支持系统中的应用[J].医学信息学杂志 ,2017,38(10):57-61,83. DOI:10.3969/j.issn.1673-6036.2017.10.013.
[21]胡兆芹, 张士靖. 领域本体的构建实践[J]. 医学信息学杂志, 2008, 29(1):12-16.
[22]张庆,吕少妮,轩扬.本体在生物医学领域中应用研究热点分析[J].医学信息学杂志 ,2019,40(1):63-67. DOI:10.3969/j.issn.1673-6036.2019.01.014.
[23]张娟.基于本体的单元信息知识组织体系构建[J].图书馆工作与研究,2017,0(12).
[24]苟梦野.基于主题词的乳腺增生病知识组织体系构建研究[D].重庆:重庆医科大学,2018.
[25]曹倩 赵一鸣.知识图谱的技术实现流程及相关应用[J].情报理论与实践,2015,38(12).
[26]黄恒琪,于娟,廖晓, 等.知识图谱研究综述[J].计算机系统应用,2019,28(6):1-12. DOI:10.15888/j.cnki.csa.00691

bmr.202003.00051V1