

基于知识组织的诊疗资源融合与发现平台设计及实现^{*}

邓盼盼 李军莲 陈 颖 夏光辉

(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〕 目的/意义 解决诊疗资源类型有限、揭示深度不足、融合关联度不高等问题，为多源异构诊疗资源的语义融合与发现奠定基础。方法/过程 提出基于知识体系的资源汇聚、基于增强表示的数据关联、基于实体对齐的内容整合等融合方法，实现专题型资源、文献型资源、公众健康教育资源的深度规范描述、语义融合与关联。结果/结论 研发诊疗资源融合与发现平台，通过智能检索与推荐、结果聚类与分面分级、语义关联等服务，支撑诊疗资源的多维度融合与发现。

〔关键词〕 知识组织；资源融合；知识发现；智能推荐；决策支持

〔中图分类号〕 R-058 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2023.11.013

Design and Implementation of a Diagnosis and Treatment Resource Fusion and Discovery Platform Based on Knowledge Organization

DENG Panpan, LI Junlian, CHEN Ying, XIA Guanghui

Institute of Medical Information, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100020, China

〔Abstract〕 **Purpose/Significance** To solve the problems of limited types, insufficient depth of disclosure and low fusion correlation of diagnosis and treatment resources, so as to lay the foundation for semantic fusion and discovery of multi-source heterogeneous diagnosis and treatment resources. **Method/Process** The paper proposes some fusion methods, such as resource aggregation based on knowledge system, data association based on enhanced knowledge representation, content integration based on entity alignment, etc., which achieves deep standardized description, semantic integration and correlation of thematic resources, literature resources and public health education resources. **Result/Conclusion** A diagnosis and treatment resource fusion and discovery platform has been developed to support the multi-dimensional fusion and discovery of diagnosis and treatment resources through services such as intelligent retrieval and recommendation, result clustering and faceted classification, semantic association, etc.

〔Keywords〕 knowledge organization; resource fusion; knowledge discovery; intelligent recommendation; decision support

〔修回日期〕 2023-06-02

〔作者简介〕 邓盼盼，助理研究员，发表论文10余篇；通信作者：李军莲，研究馆员。

〔基金项目〕 中国医学科学院医学与健康科技创新工程重大协同创新项目（项目编号：2021-I2M-1-033）；国家科技图书文献中心专项任务（项目编号：2022XM20）。

1 引言

随着健康中国建设持续推进,诊疗知识内容日渐复杂、分科更细、专业化程度更高且更新迅速,临床医生必须保持知识快速更新,才能满足日益增长的医疗保健需求,保证诊疗过程的可解释性。此时,单纯的期刊文献或专题型资源已无法满足医教科研人员对诊疗信息“精”“准”“全”的需求。而多源异构的诊疗资源分散在不同系统中单独提供服务,描述标准、标注术语和揭示深度不一致,资源充分整合与管理难度大。因此,诊疗资源的深层次融合与关联是大数据时代知识发现服务亟待解决的问题^[1-2]。本研究探索运用知识组织的理念和方法建设诊疗资源的知识体系,强化资源特征揭示和知识关联,融合汇聚结构化专题型资源及特色医学文献资源,建设诊疗资源融合与发现平台,为医教科研人员提供统一的诊疗信息全局视图。

2 资源融合现状

2.1 资源融合方式

目前主要通过信息门户、集成平台、统一元数据、资源唯一标识、关联数据等方式实现科技资源不同程度的集成。例如,为多个机构的数据库建立统一信息门户,实现网站路径的简单整合;为多个异构数据库建立统一的导航和检索界面,实现前端检索路径和结果展示的服务集成,如联机公共目录检索系统(online public access catalog, OPAC);采用统一元数据揭示资源,以基本元素为主,揭示共性特征,无法描述领域文献独有特征;基于数字对象唯一标识符(digital object unique identifier, DOI)、统一资源定位器(open uniform resource locators, OpenURL)等赋予数字资源标识,实现资源层面的标准化管理和解析定位。但上述方式均无法从疾病、药物等内容层面实现融合与关联^[3]。因此,如何实现学科领域、元数据、主题内容、服务等多个层面的资源汇聚与关联,是目前要解决的关键问题。

2.2 资源融合问题与需求

2.2.1 资源类型仍需丰富与拓展 目前医学资源主要通过文献库、知识库等形式开展服务。医学文献是医药科技创新经验和技能的积累与转化,可辅助临床科研获取最佳证据,但其规模庞大、出版周期长、质量参差不齐,难以快速凝练精华、获取最新知识。医学知识库以专家撰写的疾病、药物专题为主,更新速度快、权威性高,如临床顾问(UpToDate)^[4-5]、最佳临床实践(Best Practice, BP)^[6]等,但其关联的其他类型资源有限。因此,为应对复杂场景下的知识需求,资源类型应不断丰富与拓展,纳入症状、检查、手术等专题型资源及诊疗指南、专家共识、病例报告等特色文献。

2.2.2 内容语义表示仍需增强与细化 各类资源结构复杂、描述标准不一、关联难度大,且文献通用的元数据格式无法充分揭示诊疗指南、病例报告、患者教育等医学文献的独特内容特征。因此,为实现多类型资源的有效整合,应在通用元数据基础上进行语义增强,细化疾病、症状、药物、检查、手术等资源内容的表示维度^[7-8],为资源融合提供多维关联框架。

2.2.3 资源融合与关联度仍需提高 随着现代医学体系的迅速发展,其学科细化程度越来越高,疾病诊疗趋于专病化、复杂化。在制订诊疗方案时,为了避免片面性,除依据单科诊疗经验外,还需要多个专科或亚专科的知识支撑。但是医生较难做到对其他横向科室知识广泛涉猎、较深研究且快速整合。因此,医学资源间的汇聚与关联尤其重要,是深化知识发现服务的基础。

3 基于知识组织的诊疗资源融合与发现

3.1 资源融合与发现思路

丰富与拓展资源类型,融合疾病、化学药物、中成药、症状、检查、手术操作等专题型资源、公众健康教育资源,以及诊疗指南、专家共识、病例报告等特色医学文献资源。结合典型知识组织系统,组织专家协同建立资源的知识体系,通过专题

分类、临床专科分类、等级体系等顶层架构，纵向体现学科、同类泛指和专指资源间的层级关系；建立资源临床专科分类及其与各专题分类间的映射关系，横向聚合同领域的不同类型资源，实现学科领域层面的资源汇聚。分别建立专题型资源分级知识表示模型，对特色文献资源进行增强表示，基于丰富的关系实现资源元数据层面的规范表达与语义关

联，形成知识网络。对专题型资源的文本内容及优质医学文献进行疾病、药物、检查等实体的标注与对齐，对文献自动分类，实现资源主题内容层面的深度整合。依托资源多维度融合结果，研发知识发现系统，支持智能检索与推荐、结果聚类与分面分级、语义互联、相关资源链接等服务，为医教研人员提供可靠信息，见图 1。

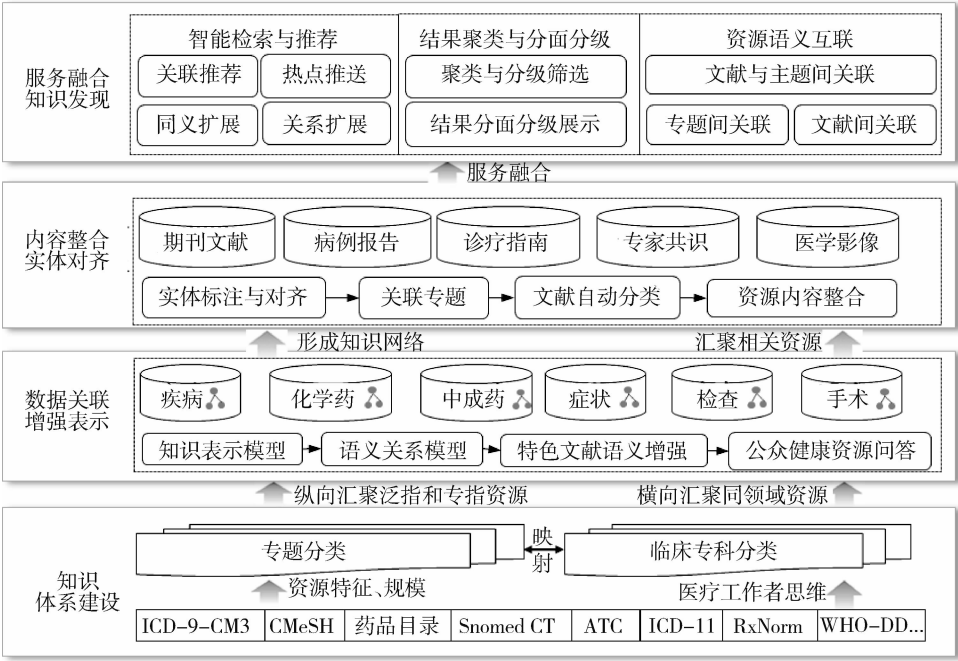


图 1 资源融合与发现思路

临床医学知识库 1.0（Clinical Medicine Knowledge Base, CMKB）建有疾病库、药物库、检查库。在此基础上，建设了 7 个专题库、3 类特色文献资源，覆盖疾病 3 000 余条、化学药物 1 200 余条、中成药 1 500 余条、症状 150 余条、检查 250 余条、手术 180 余条、患者教育 1 700 余条及相关的诊疗指南、病例报告、专家共识。同时采用 JAVA 语言完成了诊疗资源融合与发现平台的开发^[9]，以 Web 应用程序形式实现知识发现的服务融合，目前已上线试运行，面向医学生、基层医生等专业人员提供学习参考及临床诊疗所需知识，同时也为科研人员开展专病知识库建设、辅助用药、组合用药方案、预检分诊等智能医疗的探索与应用提供支撑。

3.2 知识体系建设

3.2.1 专题分类 通用文献分类与医务工作者知识分类思维差异较大，无法满足领域资源分类的精细化需求。结合各类专题型资源的个性化特征、规模分别建立专题分类，其中疾病类 29 个、药物类 32 个、症状类 32 个、检查类 34 个、手术类 28 个。国际疾病分类第 11 次修订本（11th Revision of International Classification of Diseases, ICD-11）^[10-11] 依据病因、部位、病理、临床表现等特征将疾病划分 22 个大类；中文医学主题词表（Chinese Medical Subject Headings, CMeSH）^[12] 疾病划分维度与 ICD-11 基本一致，设有 24 个下位主题类；UpToDate、BP、医景（MedScape）^[13] 等

知识库，未严格按照 ICD，而是采用 20 ~ 30 个类组织疾病专题。为了简明直观地展示和定位疾病专题资源，未采用复杂的病因病理、临床表现作为划分依据，而是综合解剖部位、医学学科和临床科室等维度，设置内分泌系统疾病、免疫系统疾病、妇产科疾病等 29 个大类组织疾病专题。世界卫生组织药物词典（World Health Organization Drug Dictionary，WHO - DD）采用解剖、治疗和化学分类系统^[14-15]将药物按照其治疗的解剖学器官/系统、治疗学、药理学和化学特点进行 5 级分类；《社会保险药品分类与代码》（LD/T 90—2012）^[16]根据药品活性物质治疗解剖系统、治疗用途及药理学作用划分西药类别，中成药则根据功能主治划分类别；《国家基本医疗保险、工伤保险和生育保险药品目录（2021 年）》^[17]结合临床实际应用情况，主要依

据临床药理学和临床科室用药分类西药，中成药则主要依据功能主治和临床科室用药分类等。考虑到与疾病专题等资源整合的需要，药物专题（包括化学药和中成药）基本采用系统部位、发挥作用、用药科室等分类依据，划分为内分泌系统用药、消化系统用药等 32 个大类；症状专题、检查专题、手术专题也主要依据系统部位、临床科室进行分类。

3.2.2 临床专科分类与专题分类映射 从医学学科角度建立资源临床专科分类，包括内分泌学、消化病学、妇产科学、神经病学、泌尿生殖学等 26 个大类。建立临床专科分类与各专题分类间的映射关系，以聚焦形成领域知识体系、横向汇聚资源。以“内分泌学”为例，汇聚内分泌相关的疾病、药物、症状、检查、手术等知识体系、专题型及文献型资源，见表 1。

表 1 专题分类与临床专科分类映射关系示例

临床专科分类	疾病专题分类	药物专题分类	症状专题分类	检查专题分类	手术专题分类
内分泌学	内分泌系统疾病	内分泌系统用药	内分泌系统症状	内分泌检查	内分泌系统手术
消化病学	消化系统和腹部疾病	消化系统用药	消化系统症状、腹部症状	胃肠道检查	消化系统手术
耳鼻咽喉头颈外科学	耳鼻喉和头颈部疾病	耳鼻咽喉科用药	耳部症状、鼻症状、咽喉症状 面部症状、颈部症状	耳鼻喉诊断 头颈部检查	耳部手术、鼻手术、咽喉手术 头颈部手术
妇产科学	妇产科疾病	妇产科用药	乳腺症状 妇科症状	乳腺检查 女性生殖系统检查	乳腺手术 女性生殖器官手术
		计划生育用药	孕产症状	孕产检查 计划生育检查	产科手术 计划生育手术

3.2.3 专题等级体系 专题型资源按照类、族、组、型等级的原理架构体系，见图 2。

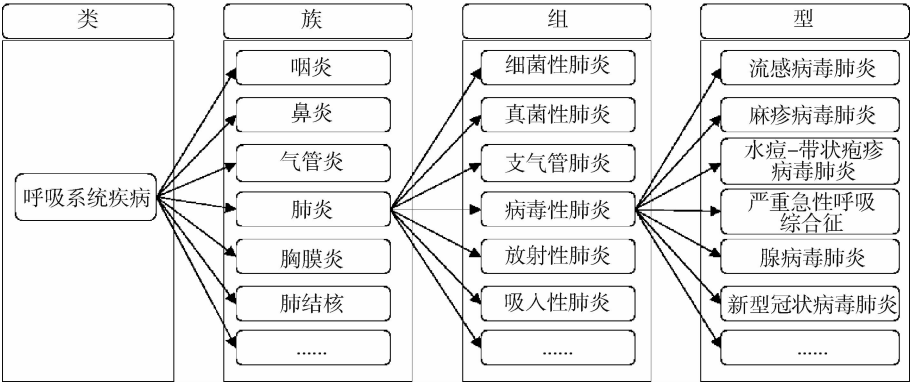


图 2 疾病专题等级体系示例

以疾病专题为例，围绕 29 个疾病专题分类，以 CMeSH 为基础梳理疾病族、疾病组及其分型，形成具有等级关系的疾病知识体系；基于中国生物医学文献服务系统（Chinese Biomedical Literature Service System, SinoMed）^[18] 统计疾病概念词频，明确疾病专题的建设范围，覆盖常见病、多发病、慢性病、重大疾病。疾病族汇聚一类具有共性特征的疾病，如“肺炎”，提供概述性通用介绍；疾病组则对应一个疾病专题，如“病毒性肺炎”，根据危急程度、病因等汇聚一组预防、诊断、治疗、随访等方案相近的分型疾病；如果分型疾病诊疗方案具有较大差异，则在疾病组下细分为分型疾病并单列疾病专题，否则在疾病组中细分疾病危重程度及诊断治疗方案。

3.3 基于增强表示的数据关联

3.3.1 知识表示模型 分级知识表示模型着力解决资源描述层次不深的问题。结合诊疗流程和健康自测真实信息需求，根据各类资源特征自定

义分级的知识表示模型，细化内容描述深度。如疾病数据模型，包括 ICD 编码、流行病学、病原学、病史、症状体征、检查、鉴别诊断、诊断标准、治疗、预防与控制、并发症等内容，重点内容进一步细化至二级描述模型，将“治疗”内容细化为治疗方案、药物治疗、手术治疗等二级元数据，细化后的描述内容也有利于“治疗化学药”“治疗中成药”“推荐手术”等关系的自动抽取；药物根据成人、孕妇、儿童、老年人等不同人群细化用法用量。

3.3.2 语义关系模型 丰富的语义关系有助于实现知识体系与各类型资源间的关联。疾病专题可以通过“鉴别诊断疾病”“并发症”“临床表现”“推荐检查”“辅助检查”“治疗化学药”“禁忌化学药”“治疗中成药”“禁忌中成药”“推荐手术”“推荐专家”“发病部位”“推荐就诊科室”等关系，关联至其他相关疾病、症状、检查、化学药、中成药、手术操作等专题型资源，为诊疗资源的融合发现服务提供支撑，见图 3。

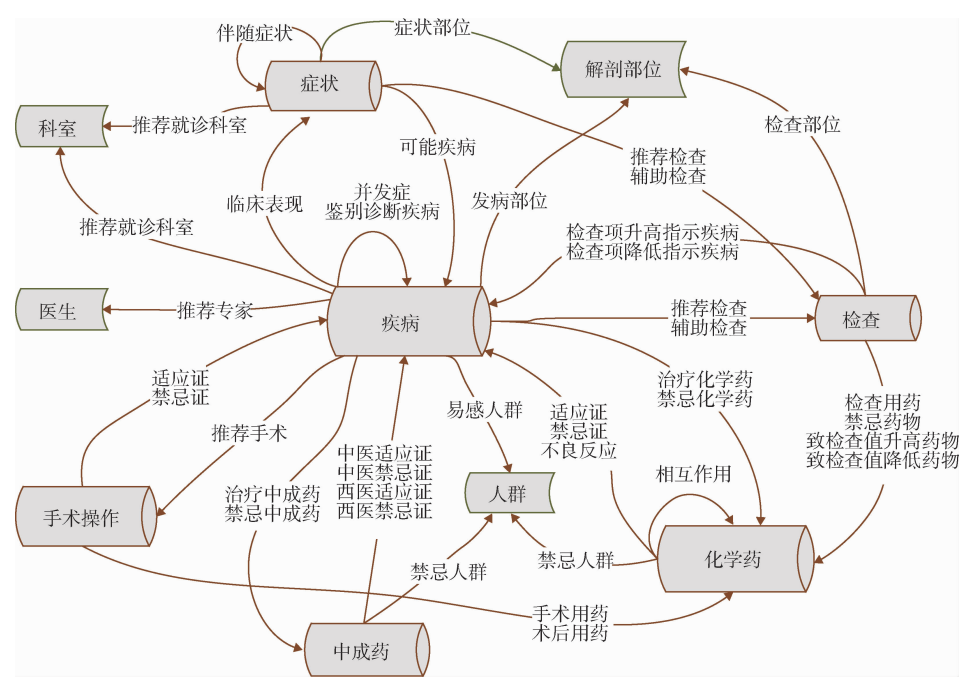


图3 专题语义关系模型

3.3.3 特色文献资源语义增强 为了突出医学特色资源内容特征，应基于通用元数据标准进行语义增强，强化文献内容特征揭示，扩增主题词、临床

专科分类、疾病、药物、症状、检查、手术、患者群体、文献类型等描述信息。

3.3.4 公众健康资源问答 公众健康教育资源继

承对应专题型资源的部分描述内容、专题分类及全部语义关系。采用结构化问答模式，由固定结构的几个问题出发，从对应专题自动获取相关问题的概要描述内容，由专家团队将专业内容以通俗易懂的语言表达出来，并通过继承的语义关系关联其他类型资源，见表 2。

表 2 “糖尿病患者须知”生成示例

结构化问题	疾病：糖尿病	对应数据项
什么是 **？	什么是糖尿病？	疾病概述
** 是怎么发生的？	糖尿病是怎么发生的？	病因
** 的症状有哪些？	糖尿病的症状有哪些？	症状
需要做哪些检查确诊 **？	需要做哪些检查确诊糖尿病？	诊断概述
怎么治疗 **？	怎么治疗糖尿病？	治疗综述
怎么预防 **？	怎么预防糖尿病？	预防与控制概述
更多帮助与支持	更多帮助与支持	-

3.4 基于实体对齐的内容整合

3.4.1 实体标注与对齐 基于位置和语句特征在特定数据项中抽取实体及关系，如在疾病“鉴别诊断”文本中识别相似疾病，标注为“鉴别诊断疾病”关系。专题型资源和文献型资源抽取的实体与知识体系中的规范词存在词形不一致情况，如“二甲双胍”药物专题“适应证”抽取实体“非胰岛素依赖型糖尿病”，无法直接关联到疾病专题“2 型糖尿病”。因此，基于 CMeSH 将主题词“糖尿病，2 型”的款目词“胰岛素依赖型糖尿病、成熟型糖尿病、慢发作型糖尿病、2 型糖尿病、成年型糖尿病、非酮症性糖尿病、非胰岛素依赖型糖尿病、稳定性糖尿病”设置为疾病专题“2 型糖尿病”的其他名称，则可实现实体名称的规范与对齐。由于词表未能覆盖全部别名，抽取的部分实体无法与现有专题对齐，如“成人发病型糖尿病”，有助于扩充知识体系的同义关系，也为拓宽专题型资源建设范围提供参考。

3.4.2 文献型资源自动分类 文献型资源经规范

标注后关联至相应专题型资源，实现自动赋类。以“非典型 1 型糖尿病 2 例报告”为例，标注的确诊疾病“1 型糖尿病”、误诊疾病“2 型糖尿病”、治疗药物“二甲双胍”专题分类分别为“内分泌系统疾病”“内分泌系统疾病”“内分泌系统用药”，基于专题分类与临床专科分类的映射关系可以为该文献自动赋类“内分泌学”。除了参考的循证文献外，围绕专题型资源，可以名称及别名扩展检索策略，从文献数据库（如 SinoMed）中有针对性地汇聚和发现特色文献资源。

3.5 基于知识发现的服务融合

3.5.1 智能检索与推荐 基于核心字段权重顺序、知识体系及语义关系等内置资源检索与推荐策略，进行检索词同义扩展、等级扩展、关联专题扩展，可实现智能检索、相关资源推荐和结果排序。如输入检索词“糖尿病”，在检索词推荐、检索结果呈现、专题信息浏览时均可获得相关资源推荐，疾病如妊娠糖尿病、2 型糖尿病、1 型糖尿病、糖尿病肾病等，药物如门冬胰岛素、精蛋白锌胰岛素、赖脯胰岛素、消渴丸等，检查如尿糖测定、胰岛素自身抗体、葡萄糖耐量试验等，症状如多尿、多食、多饮等，手术如胰腺移植术、胰腺细胞移植、干细胞移植等，辅助用户便捷获取、全面了解相关资源。

3.5.2 结果聚类与分面筛选 基于资源临床专科分类和类型进行检索结果聚类，如临床专科分类“内分泌学”下可以有效汇聚相关专题型及文献型资源。在单库检索或仅呈现某类资源时，基于增强表示的数据和关系模型进行结果分面定位，如通过“疾病名”“鉴别诊断”“并发症”等聚焦关键词出现的位置，获得名称词形相近的疾病、症状相似的鉴别诊断疾病、与之并发的其他疾病；通过“药物名称”“药物相互作用”“适应证”“禁忌证”等聚焦名称相近的药物、与该药物发生相互作用的其他药物、适用某疾病的治疗药物、某疾病禁用的药物。在检索结果页面显示资源要点内容，在资源详情页展示全面详细信息，实现资源内容分级展示。

3.5.3 资源语义互联 以“2 型糖尿病”为例，疾

病专题展示了从病因病理到诊断治疗的详细内容,实现了与各类型专题及特色文献资源的关联融合。以病例报告“非典型1型糖尿病2例报告”为例,关联了确诊疾病“1型糖尿病”、误诊疾病“2型糖尿病”、治疗药物“门冬胰岛素”,相关症状“口干”“多尿”“视物模糊”,相关检查“血常规”“尿常规”,公众健康资源“1型糖尿病患者教育”,诊疗指南“中国1型糖尿病胰岛素治疗指南”等资源。

4 结语

面向诊疗资源的集成服务需求,基于知识体系、增强表示、实体对齐等方法,实现专题型和文献型资源在学科领域、元数据、主题内容各层面的融合,研发平台支持智能检索与推荐、结果聚类与分面分级、语义互联等知识发现服务融合,为医教研人员提供一站式高质量诊疗资源。下一步将考虑融合更丰富的多模态资源,如医学影像、手术和检查视频等。

参考文献

- 1 范家巧. 大数据时代图书馆数字资源的融合与转化 [J]. 图书馆论坛, 2020, 40 (5): 38-44.
- 2 刘海鸥, 李凯, 姜波. 移动图书馆推荐系统中的用户画像与资源画像情境化融合研究 [J]. 图书馆, 2021 (6): 66-71, 93.
- 3 贾君枝. 基于关联数据的 LAM 馆藏资源整合 [J]. 晋图学刊, 2022 (3): 28-33.
- 4 Wolters Kluwer Health. UpToDate 临床顾问 [EB/OL]. [2022-11-21]. <https://www.uptodate.cn/home/about-us>.
- 5 IPEMA H J, ELSTE J M, KOPPEN L M, et al. Inclusion of optimal guidelines for pharmacists in UpToDate [J]. The annals of pharmacotherapy, 2022, 56 (9): 1084-1085.
- 6 KWAG K H, GONZÁLEZ - LORENZO M, BANZI R, et al. Providing doctors with high - quality information: an updated evaluation of web - based point - of - care information summaries [J]. Journal of medical internet research,

- 2016, 18 (1): e15.
- 7 LI P H, CHEN T F, YU J Y, et al. PubMedKB: an interactive web server for exploring biomedical entity relations in the biomedical literature [J]. Nucleic acids research, 2022, 50 (W1): W616 - W622.
- 8 SINGH A K, KHAN I R, KHAN S, et al. Multichannel CNN model for biomedical entity reorganization [EB/OL]. [2023-02-21]. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8957457/>.
- 9 中国医学科学院/北京协和医学院医学信息研究所/图书馆. 临床医学知识库 [EB/OL]. [2022-11-21]. <http://www.cmkb.cn/index>.
- 10 HARRISON J E, WEBER S, JAKOB R, et al. ICD - 11: an international classification of diseases for the twenty - first century [J]. BMC medical informatics and decision making, 2021, 21 (6): 206.
- 11 CHUTE C G, ÇELİK C. Overview of ICD - 11 architecture and structure [J]. BMC medical informatics and decision making, 2022, 21 (6): 378.
- 12 孙海霞, 李军莲, 李丹亚, 等. 基于 CMeSH 语义系统的领域自由词 - 主题词语义映射研究 [J]. 现代图书情报技术, 2013 (11): 46-51.
- 13 BYRNE G. MedScape [J]. Nursing standard, 2015, 29 (51): 17.
- 14 乔幸潮, 陈超, 李宗友, 等. RxNorm、WHODrug、SNOMED CT 三大药物术语集简介及比较研究 [J]. 中国药房, 2019, 30 (10): 1297-1301.
- 15 GALLO K, GOEDE A, PREISSNER R, et al. SuperPred 3.0: drug classification and target prediction - a machine learning approach [J]. Nucleic acids research, 2022, 50 (W1): W726 - W731.
- 16 人力资源和社会保障部. 社会保险药品分类与代码 [S]. 北京: 中国标准出版社, 2012.
- 17 国家医疗保障局, 人力资源和社会保障部. 国家基本医疗保险、工伤保险和生育保险药品目录 (2021 版) [M]. 北京: 中国劳动社会保障出版社, 2022.
- 18 中国医学科学院/北京协和医学院医学信息研究所/图书馆. 中国生物医学文献服务系统 SinoMed [EB/OL]. [2022-12-21]. <http://www.sinomed.ac.cn/>.