

DOI:10.19803/j.1672-8629.20240705 中图分类号: R917 文献标志码: A 文章编号: 1672-8629 (2024) 12-0721-09

毒性药材专题知识库构建及应用

李小阳¹, 王怡琳², 许静², 佟琳¹, 翟敏³, 蒋文博³, 何燕贤³, 雷蕾^{1#}, 游云^{2*} (¹ 中国中医科学院中医药信息研究所, 北京 100700; ² 中国中医科学院中药研究所, 100700; ³ 北斗天下卫星导航有限公司, 北京 100160)

摘要: **目的** 探讨建立毒性药材知识库, 为安全使用中药材提供参考。**方法** 研究采用知识图谱存储结构结合可视化技术组织毒性药材的毒性靶器官、毒性剂量等信息, 搭建毒性药材可视化平台。**结果** 研究构建的毒性药材专题知识库对毒性药材、含有毒性药材的中成药和毒性成分的毒性实验报道分别进行系统汇总, 实现毒性药材知识整合和数据关联性的提高。**结论** 本平台将为有毒中药的应用研究、发展和传承提供数据服务支撑。

关键词: 毒性药材; 知识库构建; 知识图谱; 药品不良反应; 毒性剂量

Construction and Application of Knowledge Base for Toxic Herbal Materials

LI Xiaoyang¹, WANG Yilin², XU Jing², TONG Lin¹, ZHAI Min³, JIANG Wenbo³, HE Yanxian³, LEI Lei^{1#}, YOU Yun^{2*} (¹Institute of Information on Traditional Chinese Medicine, China Academy of Chinese Medical Sciences, Beijing 100700, China; ²Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing 100700, China; ³BeiDou Satellite Global Navigation Co., LTD, Beijing 100160, China)

Abstract: Objective To establish a knowledge base of toxic traditional Chinese medicine. **Methods** A knowledge graph storage structure combined with visualization technology was adopted to present the information about the toxin-targeted organs and toxic doses. A visualization platform was built for toxic medicinal materials. **Results** The knowledge base exclusively related to toxic traditional Chinese medicine constructed by the study summarized the reports about toxic experiments on toxic traditional Chinese medicine, traditional Chinese patent medicines and simple preparations containing toxic traditional Chinese medicine and toxic ingredients, integrated the knowledge of toxic traditional Chinese medicine and kept data more correlated. **Conclusion** This platform will provide data service support for the applicational research, development, and inheritance of toxic traditional Chinese medicine.

Keywords: Toxic Traditional Chinese Medicine; Knowledge Base Construction; Knowledge Graph; Adverse Drug Reaction; Toxic Dose

中医药典籍是中华民族文化的瑰宝,也是中医学守正创新的基础。古人对中药毒性的认识由来已久,在本草古籍中也有不同记载。随着中医药学的发展,公众对中药材毒性的认识不断深入,历版《中华人民共和国药典》对药材毒性的描述也在不断更新、完善。据统计,《中华人民共和国药典》(2020年版)(简称“《中国药典》”)一部^[1]中收录的毒性药材已达 83 种^[2]。在临床应用以及中药新药研发过程中,确保这些毒性药材的安全有效和质量可控是当前中医药发展道路上的重中之重。随着大数据时代到

来,高通量广覆盖的组学深入研究,为中药毒性药材提供了越来越多的信息。至今为止国内外已有各种毒性领域研究类型的数据库,如药物性肝损伤专业数据库(Liver Tox)、毒物基因组学数据库(Comparative Toxicogenomics Database, CTD)、副作用资源数据库(Side Effect Resource, SIDER)等^[3-5],本研究以《中国药典》中收录的毒性药材为基础,连同国务院发布的《医疗用毒性药品管理办法》中 27 味毒性中药材品种,建立毒性药材专题知识库,将古籍文献中对毒性中药的毒性描述记载与毒性药材现代研究的毒性剂量、毒性类型、毒性表现、毒性靶器官等全部毒性信息整合,并引入知识图谱技术探索毒性药材及其毒性物质基础,通过整合多层次、多维度(古今中外)的信息,形成了毒性药材的毒性全链条知识结构,作为文献数据服务于科研工作,也为监管部门提供决策依据参考。

基金项目: 国家自然科学基金资助项目(82274244); 中国中医科学院科技创新工程(毒性药材专题知识库构建及应用 CI2021A03709)。

作者简介: 李小阳,女,硕士,助理研究员,中医药信息学。

***通信作者:** 游云,女,博士,研究员,中药药理学。

E-mail: yyou@icmm.ac.cn

为共同通信作者。

1 数据库构建

1.1 数据来源

毒性药材数据根据《中国药典》^[1]和国务院颁布的《医疗用毒性药材管理办法》(中华人民共和国国务院令 第 23 号),将中药按毒性分为大毒、有毒和小毒三类。中药的性味归经、基源、禁忌等特性来源于《中国药典》^[1]和《中华本草》^[6]。

本草数据,基于中医古籍文献知识库、国医典藏数据库、中华医典等图文数据库进行文献检索,收集历代古籍中有毒药材的记载。查阅古籍原文进行内容审校,同一种古籍的不同版本以较早版本为准。书目纳排原则:①纳入 1911 年及以前著述出版的中医药学文献;②纳入本草类古籍专著;③排除丛书与综合类古籍中记载本草书目。将纳入排除后的所有数据收录为本知识库的本草数据。

中药的化学成分数据来源于 NPASS 数据库^[7],并从 PubChem 数据库补充化学成分的分子量、分子式和分子结构等信息。

中成药数据来源于 TCM Miner 中成药数据库,该数据库收录了国家药品监督管理局发布的国家中成药数据,著录项包括:药品编码、处方、功效、主治、禁忌等。

中药及中成药的毒性数据,主要来源于中医药科

技期刊分级目录中 T1、T2 级期刊文献的中药或中成药的毒性试验研究和不良反应报道,借助 TCM Miner 标注系统模块对文献进行人工标注^[8],摘录包括实验方法、毒性剂量、毒性反应、毒性指标等数据。

化学成分的毒性数据,主要来源于 PubChem 数据库中收录的 The National Institute for Occupational Safety and Health (NIOSH) 数据库和 ChemIDplus 数据库。NIOSH 和 ChemIDplus 数据库中包含大量化学物质或混合物相关的安全信息和危险数据。

1.2 数据融合

知识融合是在同一框架下分离、处理和集成来自不同数据源的异构和多样化知识,以实现数据、信息和其他视角的融合。参考《中国药典》和 PubChem 数据库等,进行实体对齐,完成术语标准化。如中药以中药的拉丁名称对齐,化学成分以 PubChem CID 对齐,将来自国内外的数据,通过统一标准融合、转化为标准的结构化数据。

1.3 数据库的建立

基于融合后的数据建立中药、中成药、化学成分和毒性数据之间相互关联,最终实现以中药毒性药材为核心的知识图谱。所有实体的属性和实例见表 1,知识库的 Schema 图见附录:图 1。

表 1 毒性药材知识图谱实体类型

Table 1 Entity Types in the Knowledge Graph of Toxic Traditional Chinese Medicine

实体类型	数据来源	实例
毒性药材	《中国药典》(2020 年版)、《中华本草》	川乌、马钱子、斑蝥等
本草	中医古籍文献知识库、国医典藏数据库、《中华医典》	川乌(古代)、马钱子(古代)、番木鳖(古代)等
中成药	TCM Miner 中成药数据库	腰痛宁胶囊、六神丸、梅花点舌丹等
化学成分	NPASS、PubChem、文献	Aristolochic Acid II、Daphnoretin、Dihydrodaphnodorin B 等
毒性实验	NOISH、ChemIDplus、文献	中药毒性实验、中成药毒性实验、化学成分毒性实验
毒性靶器官	NOISH、ChemIDplus、文献	肝脏、心脏、肾脏等
毒性剂量	NOISH、ChemIDplus、文献	最大耐受剂量(MTD)、最小致死量(MLD)、最小可见损害作用水平(LOAEL)、半数致死量(LD ₅₀)、半致死浓度(LC ₅₀)、半抑制浓度(IC ₅₀)
毒性类型	NOISH、ChemIDplus、文献	长期毒性、急性毒性、生殖毒性等
毒性反应	文献	可逆、不可逆
毒性表现	文献	过敏、发热、咽喉肿胀、呕吐等
古代文献	中医古籍文献知识库、国医典藏数据库、《中华医典》	《千金翼方》《本草汇》《新修本草》等
不良反应报道	文献	中药不良反应报道、中成药不良反应报道
性味	《中国药典》、中华本草、古籍	辛、苦、辣、沉、暖等
归经	《中国药典》、中华本草、古籍	心经、肝经、三阴经、阳明经、少阴经等
毒性	《中国药典》、中华本草、古籍	大毒、有毒、小毒、无毒等
禁忌	《中国药典》	孕妇禁用、肾病患者慎用、高血压患者忌用等
生物实验	PubChem	BioAssay id、活性剂量等
靶点	PubChem	CDK2、ACHE、CASP7 等

2 系统的构建

2.1 系统总体设计

系统遵循现有系统架构设计理念,基于组件插件事式、多层、多模块结构进行设计,采用SOA架构、J2EE技术体系、B/S应用模式,遵循GB/T 19001-2016质量管理体系的规范和标准,按照严密的数据管理策略和安全机制保障进行设计,总体架构设计充分考虑对未来技术和业务的发展适应性。系统的前端采用AntV-G6的图谱展示处理技术,为分析和可视化提供了基本功能集,并具有丰富的插件可以进行后续的扩展。后端采用JAVA编写各个功能组件的模块,这些模块负责原始数据的获取、数据库更新和维护定时任务。数据库采用国产的百度HugeGraph图数据库,使用Apache TinkerPop3图形查询语言Gremlin,用于创建图的实体、修改实体内部属性、查询等操作。系统构架图如附录:图2所示。

2.2 系统主要功能

毒性药材专题知识库包含4个服务功能模块:知识结构模块、知识管理模块、知识探索模块、算法应用模块。①知识结构模块:管理知识库的实体和关系的结构、种类,并根据用户需求修改各实体和关系的主键属性和在知识图谱中的展示属性。②知识管理模块主要包括后台数据的查询与管理功能。以列表的形式展示平台全部的毒性药材、中成药、化学成分和毒性实验等实体数据和属性数据,可根据需求对数据增删改查,有助于实现后台内容维护及后续数据的持续更新。通过输入的属性值也可以快速查询对应的特定实体。③知识探索模块:是知识图谱的展示模块,可以根据前端输入的指令,在知识图谱中展示中药、中成药、化学成分和毒性实验实体和实体之间的关系。平台的交互功能可以帮助用户更详细地查看相关数据,具体表现为:用户通过单击知识探索模块中的实体节点,可查看实体属性详情信息。通过双击实体节点,用户可以自定义或选择调整与该实体节点相关的实体类别和每级扩展节点个数,生成二级关联图谱(功能支持生成多级关联图谱)。此外,在该功能模块二级分类下可以列表形式呈现被搜索相关实体,和所有该实体拥有的关系属性,方便用户查看。如可在“中成药”标签中查询“配伍组成”属性包含川乌的数据。④算法应用模块,平台内置了最短路径和全路径2种算法,用户可以通过简便快捷的操作,快速实现查找2个特定节点之间的最短路径查询与全部路径查询。平

台采用HugeGraph双向广度优先搜索(Bidirectional Breadth-First Grid Planning)算法。

3 平台应用与案例

3.1 毒性药材的急性毒性

根据实体查询可以检索到本平台收录的已发表文献中符合检索条件的全部毒性药材实验,对LD₅₀值进行分段查询,且单击实验节点还可查看具体实验条件和动物信息。知识图谱直观展示毒性药材各类提取物对于动物的急性毒性试验结果。根据世界卫生组织(WHO)对于急性毒性的分级,LD₅₀值大于5 000 mg·kg⁻¹的供试品被视为实际无毒,501~5 000 mg·kg⁻¹被视为低毒,51~500 mg·kg⁻¹被视为中等毒,小于50 mg·kg⁻¹被视为剧毒甚至极毒。在知识库中分级检索(附录:图3),图谱显示虽然斑蝥是大毒药材、蟾酥是有毒药材,但其粉剂灌胃小鼠的LD₅₀分别为137、350 mg·kg⁻¹,按照WHO急性毒性分级归为中等毒性;关木通中由于分离检测出的肾毒性成分马兜铃酸而被广泛关注,但其药材水煎剂的LD₅₀远大于5 000 mg·kg⁻¹,使用时应重点关注其长期毒性;川乌、草乌、马钱子的实验研究结论差异极大,跨越剧毒到实际无毒的分级,查阅者可以通过溯源原始文献,进一步明确其差异来源。同时上述数据也提示中药毒性反应的复杂性。

3.2 大毒药材的用药禁忌

查询毒性为“大毒”到禁忌为“孕妇禁用”的最短路径(附录:图4)。结果显示《中国药典》^[1]和《医疗用毒性药材管理办法》中的所有大毒药材都为孕妇禁用的中药,与文献统计结果一致^[9]。

3.3 “孕妇禁用”药材的“生殖毒性”

以禁忌“孕妇禁用”到毒性类型“生殖毒性”进行全路径查询(附录:图5),检索发现已有发表毒理学研究证实孕妇禁用的毒性药材包含部分化学成分具有生殖毒性,从现代科学研究角度进一步佐证毒性药材禁忌定义为“孕妇禁用”的合理性。

3.4 药材毒性的认知演变

毒性,与四气五味一样,是中药药性理论的重要内容之一,是历代医药学家在发现中药、认识中药、研究中药、利用中药过程中总结出的一种药物特性^[10]。在不同的历史阶段,不同学者对中药是否有毒也往往认识不一。如《神农本草经》将药物分为三品,根据该书的编排原则,上品当为补益无毒之品,后世很快认识到其中的一些矿物药如丹沙等显然为有毒之品,所谓的“延年”和“轻”身之效,也只是中

毒反应。再如细辛, 古代本草如《神农本草经》《名医别录》《本草纲目》等文献中均未有毒性之载, 直到明张景岳的《本草正》方记载有“小毒”^[11]。本平台收录了历代的古籍记载, 可以全面展示古籍中有毒药材的毒性和药性归属, 研究以马钱子为例, 通过在图谱中对药材古代与现代记载的展示, 直观地梳理了从古至今对有毒药材毒性和药性认知的历史沿革与现状(附录: 图 6)。

3.5 毒效关系探索

有毒中药的毒效关系十分复杂。现代研究表明, 中药的有毒物质包括 2 个方面: ①非有效成分, 如白果、千里光含有无治疗作用的有毒成分, 通过合理炮制和配伍可以去除毒性防止中毒; ②既是有效成分, 又是有毒成分, 以其毒性来治疗疾病, 如川乌、草乌、雪上一枝蒿等所含的乌头碱, 若将其去掉则药效丧失。含马兜铃酸中药的安全性是社会广泛关注的热点问题^[12], 虽然一些含马兜铃酸含量高的品种如关木通、广防己、青木香等已经禁止使用, 但仍有些马兜铃科中药如细辛还在临床使用^[13]。目前还有部分中药的毒效关系尚未明确^[14], 本平台参考学者提出“毒与效”复杂互作网络的研究思路^[15], 以关木通中马兜铃酸类化合物为例, 在知识库中检索, 发现马兜铃酸 I 的毒性靶器官为肾脏和肺脏。结合 WHO 急性毒性分级标准, 提示静脉注射以及腹腔注射给药途径马兜铃酸 I 的 LD₅₀ 属于剧毒; 口服马兜铃酸 I 的毒性降低, 可被视为中等毒性。图谱展示结果提示, 当前尚未见到马兜铃酸 II、IIIa、IVa 的 LD₅₀ 报道, 但是这 3 种化合物对周期蛋白依赖性激酶 2 (CDK2, Cyclin-Dependent Kinases 2) 活性有影响, 诱导细胞周期停滞, 其活性剂量范围为 20~30 μmol·L⁻¹。通过收录的毒性实验数据和生物实验数据(通过查看实验节点详细信息进行溯源), 可视化毒-效关系网络(附录: 图 7)展示关木通“毒-效”相关的物质基础和作用靶点, 为科学认知、系统评价中药生物活性和毒性的多样性和可变性提供参考。

4 讨论

本毒性药材专题知识库集成的本草类古籍文献, 涵盖了《神农本草经》《名医别录》《新修本草》《本草经集注》《本草衍义》《本草图经》《证类本草》《本草纲目》《本草品汇精要》《本草纲目拾遗》等经典古籍资源。在对古籍数据进行数字化和结构化处理后, 还重点标注了古籍中的药材名称、性味归经、毒性、产地、采收、入药部位、炮制、形态、功效主治等

术语概念以及他们之间的关系。知识库集成的丰富古籍资源不仅支撑了研究者对于毒性药材本草考证的需求, 基于这些数据资源联合知识图谱技术, 将不同朝代、不同本草古籍和不同医家对于毒性药材的认知演变过程可视化也是本知识库的特色之一。

现代研究文献数据通过筛选高质量期刊文献人工录入和引用权威开源数据源共同组成, 在人工录入的文献中重点标注了毒性药材从制备工艺、实验动物信息(动物物种、动物体重、动物周龄、动物性别)、体外实验(组织器官、实验细胞), 到实验结果(毒性类型、毒性靶器官、毒性反应、毒性指标、毒性剂量、病理描述、毒性表现、毒性描述)等术语概念及他们之间的关系。这些数据涵盖了研究者在毒性药材现代研究中较为关注的全方位信息, 并提供了文献溯源信息, 可以有效简化研究者在文献调研时的繁琐过程。

本知识库以《中国药典》和《医疗用毒性药材管理办法》有关毒性药材的品种为切入点, 补充了引起广泛关注的中药毒性品种关木通等, 系统搜集、汇总、整理、评价中草药毒性药材古今数据信息, 形成了包含 100 余种中草药的毒性药材知识库, 实现在线编辑; 建立了本草毒性药材古籍溯源、现代研究的多层次查询及关联检索。在此基础上建立的知识库较为系统汇总了毒性药材相关知识, 可实现毒性药材古今知识的整合和数据关联, 本知识库旨在对中草药毒性药材临床应用及新药研发提供重要参考。但目前该知识库仍有不足, 如用户在使用时具有一定的操作门槛, 以及由于当前数据标注方式使数据的更新速度有所局限, 目前预训练生成聊天模型(ChatGPT)为代表的大语言模型已在自然语言处理领域取得的显著突破^[16], 相信会在未来对知识库的自动化构建会提供较大帮助。下一步本知识库系统也将继续采取持续更新机制, 逐步增加舆论关注的热点毒性药材研究和其不良反应现代文献报道。

附录 (图 1~7)

本文补充数据可在线查阅: <https://doi.org/10.19803/j.1672-8629.20240705>。

参考文献

- [1] National Pharmacopoeia Commission. Pharmacopoeia of the People's Republic of China. Volume I (中华人民共和国药典一部) [M]. Beijing: China Medical Science Press, 2020.
- [2] XU L, YAN DM, NIE HY, et al. Reflection and Exploration on the Toxicity Grading Standards of Traditional Chinese Medicine Based on the Current Status of Toxic Medicinal Materials Standards in the

- 2020 Edition of the Chinese Pharmacopoeia[J]. Journal of Chinese Medicinal Materials(中药材), 2021, 44(5): 1041–1045.
- [3] HOOFNAGLE JH, SERRANO J, KNOBEN JE, et al. LiverTox: A Website on Drug-Induced Liver Injury[J]. Hepatology, 2013, 57(3): 873–874.
- [4] DAVIS AP, WIEGERS TC, JOHNSON RJ, et al. Comparative Toxicogenomics Database(CTD): Update 2023[J]. Nucleic Acids Res, 2023, 51(D1): D1257–D1262.
- [5] KUHN M, LETUNIC I, JENSEN LJ, et al. The SIDER Database of Drugs and Side Effects[J]. Nucleic Acids Res, 2016, 44(D1): D1075–D1079.
- [6] National Pharmacopoeia Commission. Chinese Materia Medica(中华本草)[M]. Shanghai: Shanghai Scientific & Technical Publishers, 1999.
- [7] ZHAO H, YANG Y, WANG SQ, et al. NPASS Database Update 2023: Quantitative Natural Product Activity and Species Source Database for Biomedical Research[J]. Nucleic Acids Res, 2023, 51(D1): D621–D628.
- [8] WANG X, LI HY, KANG L, et al. Construction and Application of TCM Miner[J]. Chinese Journal of Library and Information Science for Traditional Chinese Medicine(中国中医药图书情报杂志), 2021, 45(4): 1–6.
- [9] YOU Y, HUANG FH, HAN L. Advances in Reproduction Toxicity Study on Traditional Chinese Materia Medica Recorded in Chinese Pharmacopoeia as Prohibited or Used with Caution During Pregnancy[J]. Chinese Journal of Pharmacology and Toxicology(中国药理学与毒理学杂志), 2018, 32(5): 364–370.
- [10] CHEN RS. Reflection and Evolution of Materia Medica from the Perspective of Understanding and Handling the Toxicity of Chinese Drugs[J]. Journal of Basic Chinese Medicine(中国中医基础医学杂志), 2024, 30(1): 104–107.
- [11] CHEN RS, ZHOU YQ. Analysis of Relativity and Variability of Chinese Herbalism Toxicity[J]. Journal of Basic Chinese Medicine(中国中医基础医学杂志), 2012, 18(11): 1257–1259.
- [12] YANG YL, WANG YL, GUI XJ. Retrospective Analysis of Rational Use of Aristolochia Manshuriensis[J]. Chinese Journal of Modern Applied Pharmacy(中国现代应用药学), 2020, 37(23): 2925–2931.
- [13] XIAN Z. Study on the Toxicity Difference and Mechanism of Representative Aristolochic Acids (AA-I, AA-IVa) in Commonly Used Aristolochiaceae Plants[D]. Beijing: China Academy of Chinese Medical Sciences, 2022.
- [14] ZHAO CJ, JIA Z, TANG Y, et al. Exploring the Toxicity Characteristics of Traditional Chinese Medicine Based on The “Toxicity–Effect” Relationship[J]. Global Traditional Chinese Medicine(环球中医药), 2024, 17(8): 1602–1606.
- [15] PAN XQ, PENG C. Research Methods and Thoughts on the Integrated Study of the Complex Interaction Network Between “Toxicity and Efficacy” Induced by The Toxic Chinese Materia Medica[J]. China Journal of Traditional Chinese Medicine and Pharmacy(中华中医药杂志), 2023, 38(4): 1421–1425.
- [16] HE YH, LI M, LUO XL, et al. Research on Entity and Relation Extraction from Traditional Chinese Medicine Knowledge Graphs Based on GPTs[J]. Shanghai Journal of Traditional Chinese Medicine(上海中医药杂志), 2024, 58(8): 1–6.

(收稿日期: 2024-09-11 编辑: 孔繁瑶)