

人机互动的中药信息系统设计

赖新梅, 周常恩, 杨雪梅*, 陈梅妹

福建中医药大学 中医证研究基地, 福建 福州 350003

摘要: 信息化时代, 科学数据成为战略资源。中医药文化源远流长, 在长期实践中积累了大量数据, 而运用现代数据库技术架起中药宝库和现代科学的桥梁, 对推动中医药现代化具有重要的现实意义。提出人机互动式中药信息系统的构建, 设计构建由方剂、中药、物种、化学成分等组成相互关联的中药信息系统, 以适应现代科研、教学的需求, 特别要满足数据挖掘(KDD)、计算机辅助药物设计(CADD)等技术在中医药领域的应用, 并详细阐述了系统框架及人机互动式个性化信息服务的查询功能。

关键词: 中药; 中医药; 信息系统; 数据库; 人机互动; 数据挖掘

中图分类号: R28 **文献标志码:** A **文章编号:** 0253-2670(2011)10-2150-04

Design of human-computer interactive information system of Chinese materia medica

LAI Xin-mei, ZHOU Chang-en, YANG Xue-mei, CHEN Mei-mei

Research Base of TCM Syndrome, Fujian University of Traditional Chinese Medicine, Fuzhou 350003, China

Key words: Chinese materia medica (CMM); traditional Chinese medicine (TCM); information system; database; human-computer interaction; knowledge discovery in database

21 世纪, 信息系统无疑将对中药现代化发挥巨大的作用。中药信息系统的研究主要是利用计算机技术、信息处理技术, 特别是数据挖掘(knowledge discovery in database, KDD)技术^[1]、数据库技术和现代的虚拟技术, 结合现代的化学元素、数学、生物学、中医药学等学科的研究成果, 建立和完善中药方剂数据库、中药化学数据库、中医药知识库等^[2]。而实现中药现代化、国际化的一条重要途径, 就是加快中药信息、知识的传播和更新速度。通过建立中药信息系统, 利用国际互联网进一步扩大中药电子信息数据的使用范围, 则是加快中药知识传播和更新速度的重要途径^[3]。因此, 本文提出一个融合传统中药及现代研究有机关联的中药信息系统, 以适应现代科研、教学的需求, 并详细阐述了系统框架及人机互动式个性化信息服务的查询功能。

1 中药信息系统构建的意义

我国中药数据库建设是从 19 世纪 80 年代开始发展起来的, 经过 20 多年的发展, 已取得初步的成

果。到目前为止, 我国已有近百个规模不同的中药数据库完成建设并投入使用, 初步实现了中药信息数字化^[4]。其中主要有中医药期刊类数据库、中医药事实型数据库及中医药文献-事实型数据库等^[4-5]。近年来, 中药数据库建设已向综合性、智能化、个性化趋势发展, 为现代中医药教学、科研提供必备的工具。上海中药创新研究中心研建的 SIRC/TCM 智能化中医药信息系统, 是一个由疾病、方剂、中药、中药化合物等多个各自独立又相互交叉连接的数据库组成的综合信息系统, 系统框架如图 1 所示。中药药用植物化学成分结构及药理活性数据库是以“药物分子设计方法研究”为背景开发而来的, 可作为中药现代化研究工具以及从中药资源出发创制新药的开发工具。其以“植物”作为关联传统中医药与现代研究的桥梁, 收集了大量在外文杂志上公开发表的, 分属天然产物、药用植物、植物药理等不同领域的研究成果, 把有关的化学成分、结构信息和药理数据“引入”了中药化学数据库, 使得该数据库不仅包括我国科学家的研究成果, 而且涵盖了

收稿日期: 2011-03-07

基金项目: 福建省科技厅重点项目(2009Y0030); 福建省自然科学(青年)基金(009J05074); 福建省卫生厅重点项目(wzzsj0901, wzzsj0902)

作者简介: 赖新梅(1977—), 女, 福建三明人, 馆员, 硕士, 从事中医药数据处理研究, 发表论文 9 篇。Tel: (0591)22861513 E-mail: xm9709@163.com

*通讯作者 杨雪梅 Tel: (0591)22861513 E-mail: yxm_wj@sina.com

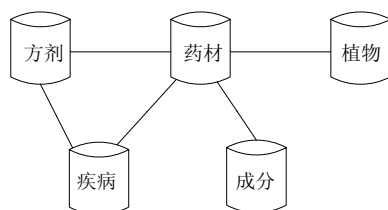


图1 SIRC/TCM 智能化中医药信息系统总框架

Fig. 1 SIRC/TCM general framework of intelligent information system of TCM

外国科学家的大量研究成果，极大地丰富了数据库的内容。

虽然 SIRC/TCM 智能化中医药信息系统，已是由疾病、方剂、中药、中药化合物等数据库组成的有机整体，但药材与成分之间直接建立关联，不利于将国内外相关化学、药理等数据引入中药信息系统；而中药药用植物化学成分结构及药理活性数据库采用了与现代植物分类学一致的建库策略，重视其他领域相关化学成分、药理等现代研究数据的收集，但未能与传统中医药知识建立有机的关联。因此，为建立完善中药数据平台，满足现代科研、教学需求，特别要适合 KDD、计算机辅助药物设计 (computer aided drug design, CADD) 等技术在中医药领域的应用，并在分子水平开展中医药的相关研究，本课题组设计开发了人机互动的中药信息系统，现就其具体框架进行简要介绍。

2 中药信息系统架构介绍

2.1 基本框架

中药信息系统是由福建中医药大学自主研发的中医药科学数据平台，包含了中药基本信息库、中药物种数据库、中药化学成分数据库、方剂数据库 4 类各自独立又相互关联的数据库。该数据库以大量信息为基础，并针对中医药信息、化学信息等各自的特点，建立了各个库之间的关联，如图 2 所示，虚线部分表示待建。中药信息系统设计特色在于化学成分、药理等现代研究数据收集与整理：分别以中药材、植物为核心收集数据，既注重收集传统中医药的现代研究数据，也重视以“植物”为桥梁，将国内外相关领域的研究成果引入中药体系。

2.2 数据蓝本的选择

方剂数据库：《中医方剂大辞典》；中药基本信息库：《中国药典》、《中华本草》、《中药大辞典》、《新编中药志》；中药物种数据库：《中国植物志》；中药化学成分数据库：《原植物化学成分手册》及国内外相关文献报道。除了上述 4 个主要数据库，中

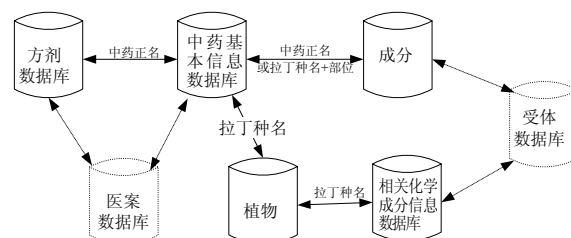


图2 中药信息系统框架

Fig. 2 Framework of CMM information system

药信息系统还建立了中国台湾中药特色数据库，收集了《“中华民国”中药典范》、《台湾原住民药用植物汇编》、《台湾青草药》、《台湾常用药用植物图鉴》、《台湾药用植物资源名录》等数据。这些特色数据目前暂时以单库的形式存在，可以进行简单的查询。方剂数据库、中药基本信息库、中药物种数据库、中药化学成分数据库之间具体关联见图 3，加下划线表示关联的关键字段。

2.3 数据处理

为了系统中各数据库的有效关联，对数据进行了一些必要的数据处理。

2.3.1 方剂“组成”结构化处理 方剂“组成”字段包含了药名、炮制、剂量等信息，且描述不完全一致。因此，中药信息系统根据“组成”字段建立“方剂组成结构化表”，包含方剂 ID、药名、炮制、用量、单位 5 个字段，并对单位进行统一换算。

2.3.2 植物物种数据的规范处理 中药基本信息库有多个数据源，且对同一中药来源物种的描述并不完全一致。因此，中药信息系统以《中国植物志》为标准，建成统一的中药物种数据库，并对物种的拉丁学名进行规范处理：“规范拉丁种名”、“规范拉丁异名”，即指将原拉丁名、拉丁异名中“命名人”的相关信息删除。由于不同的文献对拉丁名学名中“命名人”信息的描述不一致，有的是全名，有的是缩写，使得系统无法通过“原始拉丁种名”进行精确的关联，因此，中药信息系统根据拉丁名命名规则，将“命名人”信息删除，能精确建立各相关数据库的关联且不影响物种的确认；由于中药基本信息库存在拉丁种名与拉丁异名互用的情况，为了提高数据库之间关联的准确度，中药信息系统依据拉丁种名唯一性的原则，以《中国植物志》中选择的拉丁种名为“标准拉丁种名”。

3 人机互动式查询系统设计

一个优秀的中药信息系统，首先应具备强大的

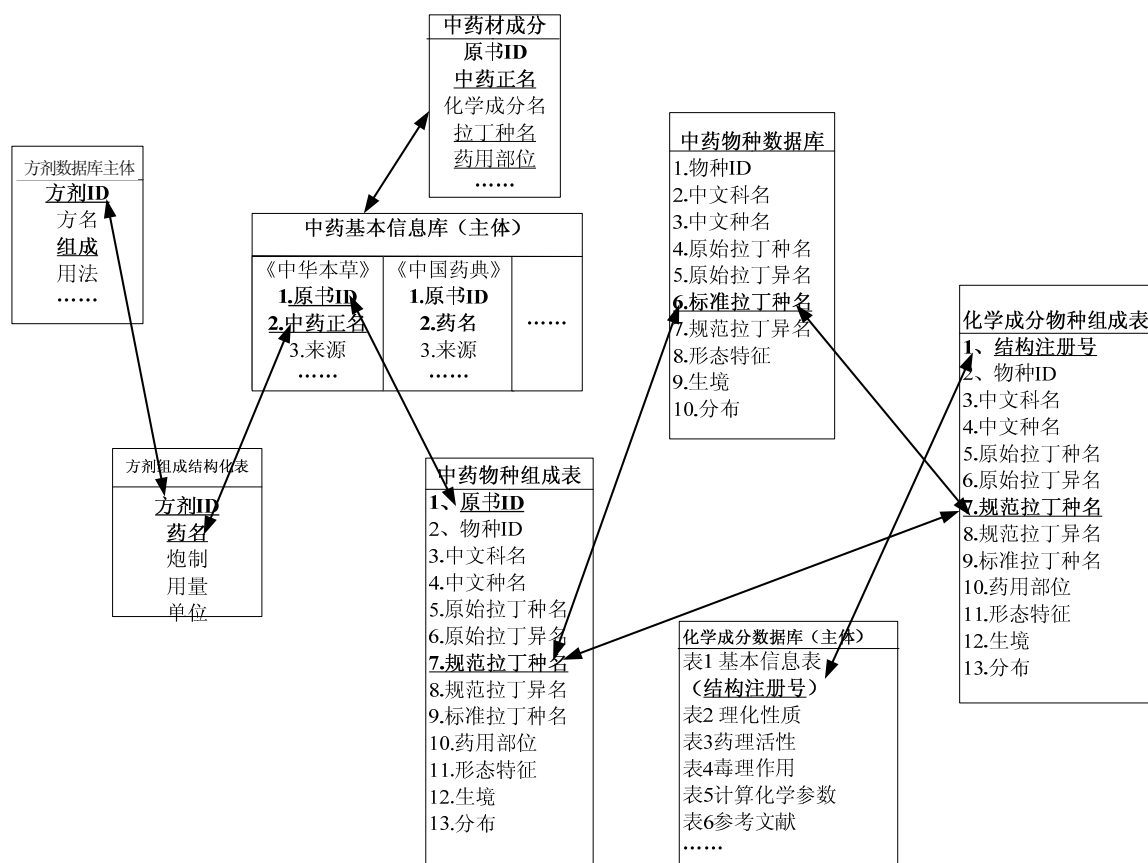


图3 中药信息系统实体关系示意图

Fig. 3 Entity relation diagram of CMM information system

查询功能, 中药信息系统既提供常规的简单查询, 又提供高级查询, 即实现跨库查询, 支持中药类、方剂类、物种类和化学成分综合库的关联查询; 历史记录查询, 即查询用户有选择保存的历史查询记录; 正则查询, 利用正则表达式来构造具有复杂文本查询条件的查询, 可以帮助用户找到符合正则表达式所定义文本模式的记录。正则表达式是一种用于描述和操纵文本数据的强大工具, 能利用各种元字符 (metacharacter) 和普通字符并遵循一定的语法规则来定义某类文本数据所遵循的模式。如查询主治感冒且在字符串“感冒”之后 11 个字符内出现“咳嗽”的记录, 其正则表达式为“感冒…………咳嗽”; 还有化学分子的二维子结构检索等。此外, 中药信息系统还设计灵活有效的人机互动式查询。

3.1 可定制查询

定制查询是指用户可自主选择系统中提供的各类数据及某类数据的具体数据源为检索入口, 并且可以对“查询结果”进行选择, 然后进一步关联到其他相关数据源。用户可以选择中药库或其他库为

检索入口, 如选择中药库时, 还可以选择单个数据源或多个数据源组合查询, 选择单个数据源时, 按该数据源的原始字段为检索入口; 当选择多个数据源时, 系统自动生成统一视图作为检索入口。查询的结果也可以定制, 如查询“麻黄”选择《中华本草》中“拉丁种名”, 用户可以查看指定物种化学成分数据, 然后依据中药信息系统本身提供的“方剂-中药-物种-化学成分”之间结构导航, 进一步进行循环导航查询。

3.2 可扩展查询

扩展查询是指用户可以选择在不同“查全率”层次对查询结果进行关联。中药信息系统主要是对“拉丁种名”进行关联的扩展。由于在整理中药信息系统数据过程中发现, 同一数据源, 如《中华本草》自身就存在拉丁种名、拉丁异名混用的情况, 因此, 与“物种”和“化学成分”关联时, 中药信息系统中默认通过“规范拉丁种名”进行关联, 用户可用两种扩展方式: (1) 拉丁种名进行关联 (规范拉丁种名或标准拉丁种名); (2) 规范拉丁异名进行关联

(规范拉丁种名或标准拉丁种名或规范拉丁异名)。通过扩展关联查询,可以尽可能地将最新化学成分领域的研究,如药理活性、毒理研究等引及中药体系,且用户可以根据自己的专业知识来判断关联的准确度。

3.3 用户自由组方查询

用户自由组方查询是指用户可以通过物种,即以“拉丁种名”来组方。中药是一个复杂的系统,中药材名与来源物种之间并非一一对应,主要有3种情况:“一对一”(一种中药材对应一个物种)、“多对一”(几种中药材对应同一物种不同的药用部位)和“一对多”(一种中药材对应多个物种)。而科研工作者在研究过程中,对于多基源中药大多是选择其中一种来源物种作为研究对象。为解决此问题,中药信息系统可选择对物种查询的结果,同时用户可自主添加“拉丁种名”,以达到用户通过“物种”自由组方,并进一步查看该方化学成分数据。

通过定制查询、扩展查询、自由组方查询等人机互动查询功能,实现中药信息系统的个性化信息查询服务,以满足不同科研工作者的需求。

4 结语

中药信息系统虽已具备一定人机互动的查询功能,并在科研中发挥了重要的作用,但还需不断更新完善,在设计思想等方面还值得深入探讨,很多地方需继续深入研究,如中药信息系统对“拉丁种名”做了扩展查询,下一阶段,将从“形态特征”、“生境”、“分布”等信息来确定一个物种,从而提高各数据库之间关联的精确度;还有一些数据库有待继续建设:如医案数据库、受体数据库等。

参考文献

- [1] 李凌艳,李认书,孙鹤.数据挖掘技术在中药研究中的应用[J].中草药,2010,41(5):附16-附18.
- [2] 肖培根,肖小河.21世纪与中药现代化[J].中国中药杂志,2000,25(2):67-70.
- [3] 陈峰,崔蒙.中药信息系统建立初探[J].中草药,2000,31(11):附1-附3.
- [4] 万仁甫,徐伟.中药数据库的现状与发展趋势探讨[J].中国药房,2006,17(10):794-796.
- [5] 李广曦,陈蔚文.中医药类数据库建设的回顾与分析[J].医学信息,2005,18(6):557-560.

《现代药物与临床》杂志征稿与征订启事

《现代药物与临床》杂志(CN12-1407/R,ISSN 1674-5515)是国家级医药科技期刊,天津市一级期刊,2009年1月由《国外医药·植物药分册》更名为《现代药物与临床》,并被美国《化学文摘》(CA)、波兰《哥白尼索引》(IC)、美国《乌里希期刊指南》(Ulrich's Periodicals Directory)、美国《剑桥科学文摘》(CSA)、英国《全球健康》(Global Health)、CNKI中国期刊全文数据库、中国核心期刊(遴选)数据库等刊载。为了进一步提高期刊质量,2010年出版的《现代药物与临床》全新改版,更加突出创新性与实用性,紧跟国内外药学发展趋势,适时追踪热点,从栏目内容、文章质量,到封面版式、装帧印刷都得到了全面的提升与改进。

办刊宗旨:报道国内外药物研究的新进展与新技术,以及药物在临床应用方面的最新动态,为新药研发、生产人员以及临床医生与药剂师合理用药提供有益的参考。

内容与栏目:涵盖药物的基础研究与临床研究各学科,设置“专论与综述”、“实验研究”、“临床研究”、“未来药物”、“药事管理”、“知识产权”、“药物经济学”和“市场信息”等栏目。“专论与综述”栏目除报道植物药研究的最新进展外,诚征有关药物与临床研究前沿的前瞻性文章。

读者对象:药物研发、生产、监管人员,以及临床医生与药剂师。

《现代药物与临床》双月刊,国内外公开发行,封面铜板彩色覆膜。为扩大信息量、缩短出版周期,本刊由64页扩版为80页;为惠顾广大读者,改版不提价,每期定价仍为15元,全年90元。本刊自办发行,请直接与编辑部联系订阅。

本刊网上在线投稿、审稿、查询系统正式开通,欢迎投稿、欢迎订阅!

《现代药物与临床》编辑部

地址:天津市南开区鞍山西道308号(300193)

电话与传真:(022) 23006823

网址:www.中草药杂志社.中国;www.tiprpress.com

邮箱:dc@tiprpress.com

开户银行:兴业银行天津南开支行 账号:44114010010081504 户名:天津中草药杂志社