

中药知识库设计浅析

王连心¹ 孟庆刚² 王志国¹ 赫炎¹ 徐璞¹

(1 中国中医科学院中医临床基础医学研究所,北京市东直门内南小街 16 号,100700; 2 北京中医药大学)

关键词 中药;知识库

中药是中医药防病治病的重要手段,在中医药学科体系中占据着重要的地位。中药往往是以方剂的形式出现的,方剂是中医医师在中医理论的指导下,辨证审因、决定治则治法,选择适当的中药,按组方原则,酌定用法用量,妥善配伍而成的治疗手段^[1]。从古至今,关于中药与方剂的著述汗牛充栋,既为方药的学习提供了浩如烟海的文献资源,与此同时,海量的信息资源也给检索与利用带来一定的困惑与不便。并且,由于中药和方剂本质上属于诊疗经验的人为总结和提炼,其整体性、辨证性、动态性等本质特征往往过于抽象化和定性化,也为其定量和精确的把握带来一定的困难^[2]。

随着信息技术与计算机技术的飞速发展,数据库技术在中医药领域也得到了广泛的应用,从以光盘为载体的数据库,逐渐过渡为以网络数据库为主,光盘数据库与网络数据库并存。机上文献相对于传统文献而言,具有节省经费、便于查询的优点^[3]。中医药数据库这种现代信息技术是在中医药信息研究过程中,针对原始资料信息量大、不规范等问题进行分析与整理、补充与完善而逐渐建立起来的^[4],并随着信息技术和计算机技术的进步而不断进步^[5]。新兴的知识库技术,又被称为人工智能数据库,可以理解为是由人工智能和数据库这两项计算机技术有机结合而成的^[6]。本文将在总结既往中药领域数据库建设特点的基础上,借鉴优点,主攻瓶颈,探讨中药知识库的构建,浅析其设计思路。

1 中药数据库简况

自上世纪八十年代,中药数据库建设在我国逐渐开展并发展,取得了一定的成果,初步实现了中药信息数字化,例如中药文献数据库、中药基本信息数据库、中药有效成分数据库、中药学图像和文字数据库等的建立与完善。总的来说,在建库技术方面,中医药类数据库随着计算机技术的进步而不断进步,中医药科研

人员接触并使用先进数据库软件的速度也在加快。就数据库的内容和表现形式而言,现有的中药数据库大体可以分为全文数据库、书目数据库和数值数据库等几种类型,如薛氏^[7]设计构建的中药信息数据系统,既包含中药的基本信息、又包含其化学及生物学表征,还可以对海量数据进行存储、管理和知识挖掘。就数据库结构而言,一方面基于个人微型计算机的桌面型数据库在不断增加,另一方面基于网络的客户-服务器结构和浏览器-服务器结构数据库也逐渐出现^[8]。王氏等^[9]开发的《中草药微量元素数据库》,是国内第一个具有图形用户界面的中医药类数据库,其出现的时间距该界面应用程序开发工具的问世还不到 5 年。在早期建库过程中,多是通过字符界面软件构建文献型库,但因其自身的缺陷往往致使数据库的应用范围有所限制,可视化编程工具的出现大大突破了字符界面软件的限制,可以设计出符合使用习惯的人机交互界面,使用户与计算机技术之间的鸿沟得到一定的填补。同时,VFP、Access 等小型数据库软件也逐渐得到采用。另外,随着网络技术的进步和信息量的增加以及信息源的扩大,基于网络的 SQL server 和 Oracle 等大型数据库软件也渐渐被大家关注。刘氏^[10]将中药科技基础信息数据库依托在一个功能强大的 Oracle 数据库系统平台上,面对其数据格式多样化,并且表与表之间的关系纷繁错杂的特点,应用 Base View 网页导航系统实现数据表之间数据相关性的良好导航,并提供精确与模糊 2 种查询,通过查询单味中药主表就可以直接查询到多个关联关系。

与此同时,数据库的形式也正发生着从单表型数据库向关联型数据库方向的变化,多库融合共享平台、数据库关联检索、数据挖掘等技术也已开始应用,面向基于互联网结构的数据库需求,还可以运用网页编程语言^[8]。虽然中药数据库建设已初具规模,但同时也存在着一些问题。据统计^[4]目前国内、外已建成的近百个中药领域数据库中,主要是以文字描述性数据库为主,另外重复建设现象偏多;再如专题数据库多,综合数据库少;全文数据库检索不便;缺乏统一的数据库标准等。李氏^[8]曾运用文献计量法,按学科主题、数据库类型,对 1984 年到 2003 年中医药类数据库进行

基金项目:中国中医科学院第三批自主选题,国家自然科学基金资助项目(No. 81072897, No. 30873461, No. 81102666)

分析,认为数据库建设的关键之一应在于借鉴先进的计算机技术。现有的中药数据库应向信息全面、更新快、资源共享的方向进一步发展完善,如建立与完善中药研究开发信息系统、完善网上共享平台、开发智能化与个性化的信息服务系统等^[4]。

智能信息等现代科学技术为此提供了技术支持,例如,数据库知识发现技术的出现在一定程度上解决了信息量大、数据复杂、不易处理的困境,并为数据库知识发现奠定了基础。所谓数据库知识发现,是指从大量数据中提取出可信的、新颖的、有效的并能被人理解的模式的过程^[11]。数据库知识发现技术的发展使从浩如烟海的中药信息中获得能够反映中医药理论隐含的、未被人们所知的规律和模式成为可能,这将大大提高中药数据库的智能化水平,便于对中药各方面知识的发现与挖掘,从而推动中药现代化进程。中药知识库就是在如此的背景与要求下发展起来的,是集计算机科学与中医药科学于一体的交叉学科信息技术,但是由于2个领域在研究方向与研究方式上存在着一定的差异,因而还需要不断的契合与融合。

2 中药知识库设计浅析

2.1 数据库与知识库比较^[6,12]

2.1.1 定义的对比 数据库是存储在一起的相关数据的集合,可谓是一种按照数据结构来组织、存储和管理数据的仓库。知识库的定义来自传统的数据库和人工智能及其分支——知识工程这2个不同的领域,是知识工程中结构化、易于操作和利用的知识集群,也是面向领域问题求解的相互关联的知识片集合。

2.1.2 特点的对比 数据库中的数据主要具有整体性和共享性的特点,其数据结构是基于数据间的自然联系,并摆脱了具体程序的制约。知识库中的知识片多为模块化的,既包括与领域相关的理论知识和事实数据,也包括启发式知识,在事实数据层的基础上还有准则层与策略层,知识片之间通常也都存在相互依赖的关系。

2.1.3 区别与联系 数据库系统具有存储海量数据的能力,但不能表达和处理规则等知识。知识库是知识工程中的知识集群,可以使基于知识的系统具有智能性,因而,可以在关系型数据库的基础上,构建知识库,规范事实数据层的知识信息,并通过准则层来实现决策层的知识发现。

构建中药领域知识库,按照知识库的基本结构应将其主要分为3个层次,第一个层次是事实数据层,包含存在于文献资料中的原始数据,诸如中药名称、出处、剂量、方剂命名、药物组成、配伍规律等相关数据;

万方数据

第二个层次是基于创建规则的知识本体,其内容是在事实数据中抽取出来的,并通过创建规则划分的不同概念,并包含于概念之中;第三个层次是基于前两个层次的策略层,是对已获得的知识得出策略性的结论和规则。

2.2 中药知识存在的问题

2.2.1 数据信息缺失与不规范 中药往往具有一药多名或重名的现象,对其基源的描述也存在一定的不一致。同一味中药在不同地区可能有不同的名称,不同中药在不同地区也可能有相同的名称^[13]。沈氏^[14]在研究中发现,即使权威水平的数据源,其数据也存在一定的缺失、错误和不规范现象,同物异名、同名异物现象较为严重。因此,在构建中药知识库的过程中,数据加工处理是不可或缺的工作。

2.2.2 界定标准难以精确和定量 中药知识是我国人民在长期医疗实践过程中的经验总结,因而许多中药信息是靠经验来描述的,也就造成了中药信息的模糊性和不确定性。例如,关于中药的毒性,往往只是被描述为有毒、无毒、大毒、小毒或剧毒等,而对各种毒性缺乏定量的描述,如半数致死量之类;再如,中药的“味”可分为甘、微甘、苦、微苦等,但由于人的味觉存在差异,同一味中药可能有人觉得是甘,而有人则觉得是微甘,其间也缺少一个定量的划分标准^[15-16],这都是中药信息模糊和不确定的表现,也是构建中药知识库需要重点解决的问题之一。

2.2.3 药效术语抽象与庞杂 中医药历史经年久远,中药相关典籍也十分庞杂,中药药效名词术语种类较多,包含了海量的数据信息,并且不同典籍关于药效名词的使用也存在较大差异,因而其成为中药信息学研究的一类重要数据类型和研究对象。同时,有些药效名词使用频率较低,对应的中药也较少,这也使药效知识的规范成为中药知识库构建需要攻克的又一个难关。

由上可见,中药知识的特殊性就在于中药理论的语言以及概念的特殊性,在数据库建设的基础上,扩大知识库建设规模和支持类型^[17],开发基于数据库的中药领域知识分析与关联挖掘系统,将是中药知识库设计与建设的目标。

2.3 中药知识库的设计

2.3.1 对象 中药知识是提高中医临床与学术水平的重要环节和文献宝库之一,值得当代医家学习与研究,若离开对本底知识的发掘与研究,学科的发展可能会逐渐失去其优势与特色^[18]。由上所述,既往的中药数据库或方剂数据库存在存储知识疏于规范、功能单一的不足。因而,中药知识库面向的对象本质上是适

于临床应用的中药知识,可通过对古典医籍、学术专著、学术期刊和报纸、各种学术会议论文和报告以及相关指南等进行普查与筛选、分析与整理而获得。

2.3.2 构建关系型数据库 构建关系型数据库,使之作为中药知识库的事实数据层,以表的形式存储在知识库系统中的事实部分。中药知识包含的信息是多方面的,既有药的,也有方的,还有诊断的,并且其间的关系也往往纵横交错。因此,对中药知识的系统存储是关键的一个环节,既应减少冗余,又要方便系统快速处理,还要有灵活性和扩展性。总体上讲,关系型数据库由基本表、预处理表、用户交互表和临时表组成^[15]。就基本表而言,创建中药表,包括正别名、异名、处方用名、基源、性味、归经、用药宜忌等;创建功效主治表,包括功效、主治、证、主症、次症及用法用量、常用配伍等;创建方剂表,包括含有某种中药的方剂以及方剂的处方来源、剂型、药物组成、功效主治、制备方法、用法用量与禁忌等,以及中药用药剂量转换表、特殊剂量表等。同时,定义每张表的字段名、字段类型、长度、注释等,并且,每张表都以“中药名”为主键,建立表之间的关系。

2.3.3 从数据库到知识库 将关系型数据库作为事实数据层,在其基础上建立准则层和决策层,使中药数据上升为中药知识,并统一规范与分门别类,构建中药知识库。纳入提取的中药和方剂相关知识点,并根据中医理论,将中药作为本体,分别确立中药与方剂知识中的类、类的层次及其关系。

3 讨论

3.1 检索查询 知识库支持用户多层次与多维度的检索查询。可以在查询框内输入检索词,单击查询按钮显示结果,通过任意一个关键词,可得到不同层次与不同维度的查询信息。

3.2 命名规范 同一味中药往往具有正名、别名、异名、释名等,在不同的方剂中也可能出现多个不同的名称。为解决一药多名的问题,基于标准化研究的简单性原则,对中药名称进行梳理整合,规范统一,在中药知识库中实现正名与异名、别名、释名之间的自动转换,并不断完善与更新。

3.3 剂量规范 中药剂量单位往往存在不统一的现象,重量单位“两”“斤”以特殊剂量单位“枚”“条”等也可能出现在同一个方剂(尤其是古代文献)中,由于现代中药计量单位统一为国际单位“克”,因此,参照《中国历代度量衡考》《中国科学技术史·度量衡卷》等,在中药知识库中通过函数实现剂量转换的功能。

万方数据

基于中药知识库,可以运用 Delphi、AHP 等方法对中药和方剂的性味、归经、功效、用药规律等知识片进行深入挖掘与关联探讨,也可以通过对中药名、主症、次症和剂量的标准化转换,将噪声数据过滤,形成规范性知识,推动中药及方剂标准化、规范化研究向前发展,这方面将在以后的研究中进行深入探索。

参考文献

- [1] 王付,等. 方剂学[M]. 北京:中国中医药出版社,2010:2-5.
- [2] 龚燕冰,倪青,王永炎. 中医证候研究的现代方法学述评. 北京中医药大学学报,2006,29(12):797-801.
- [3] 张黎. 21 世纪图书馆文献采访:论机上文献与传统文献采访的关系. 图书馆工作与研究,2000,(4):43-45.
- [4] 万仁甫,徐伟亚. 中药数据库的现状与发展趋势探讨. 中国药房,2006,17(10):794-796.
- [5] 刘海波,彭勇,肖培根. 当前中药数据库建设中的几个问题. 世界科学技术—中医药现代化,2009,11(3):339-343.
- [6] 施荣明,赵敏,孙聪. 知识工程与创新[M]. 北京:航空工业出版社,2009:20-33.
- [7] 薛兴亚,徐青,章飞芳. 中药信息数据系统构建. 世界科学技术—中医药现代化,2006,8(3):91-94.
- [8] 李广曦,陈蔚丈. 中医药类数据库建设的回顾与分析. 医学信息,2005,18(6):1-4.
- [9] 王麟生,乐美卿. 中草药微量元素数据的数据库管理. 广东微量元素科学,1996,3(1):68-70.
- [10] 刘岩. 中药科技基础信息数据库的研究与应用. 中国中医药信息杂志,2009,16(增刊):65-67.
- [11] 张春华,王阳. 数据挖掘技术、应用及发展趋势. 信息化与网络建设,2003,(4):47-50.
- [12] 李建中. 数据库系统原理[M]. 北京:电子工业出版社,2008:3-50.
- [13] 吴园园,章新友. 中药信息量子化研究. 江西中医学院学报,2008,20(2):56-57.
- [14] 沈国俊,林端宜,周常恩,等.《中药大辞典》植物药条目分析. 福建中医学院学报,2009,19(2):54-58.
- [15] 胡建军. 中药方剂数据挖掘系统中数据存储的设计. 福建电脑,2008,(6):3-4.
- [16] 胡波. 中药方剂性味的多维宏观量化表达方法研究[D]. 成都中医药大学硕士论文,2003:36-39.
- [17] 中医现代化科技发展战略研究课题组. 中医药相关数据库建设. 世界科学技术—中药现代化,2002,4(5):24-27.
- [18] 胡滨. 中医药文献检索[M]. 上海:上海科技出版社,2007:33-45.

(2011-09-14 收稿)◎

敬告作者

根据国家中医药管理局“国中医药发[2006]63号”《中医药继续教育规定》、《中医药继续教育登记办法》的通知附件:“三、中医药继续教育学计算方法”第(二)项第5条“在学术期刊上发表论文,按刊物级别授予学分”之规定,《世界中医药》在国内外公开发刊,具有国际标准连续出版物号 ISSN 1673-7202 和国内统一连续出版物号 CN 11-5529/R。凡在《世界中医药》刊登文章第一至第三作者分别授予 6、5、4 学分(余类推)。如需学分,请与《世界中医药》编辑部联系。