



本文经编委遴选,英文版将通过 ScienceDirect 全球发行。

基于领域本体方法构建 中医概念信息模型的思考*

□ 谢 琪** 崔 蒙** (中国中医科学院中医药信息研究所 北京 100700)
曹存根 (中国科学院计算技术研究所 北京 100190)
刘保廷 (中国中医科学院 北京 100700)
刘长林 (中国社会科学院哲学研究所 北京 100732)
焦宏官 (中国中医科学院中医药信息研究所 北京 100700)

摘 要:信息学方法与技术在中医药学科领域得到了日益广泛的应用和发展,由于缺乏中医概念信息模型的研究与突破,顶层设计的理论指导不足,中医药信息研究多以需求为导向,自下而上的建设,中医药的特点和优势难以体现。本文对中医概念信息模型的理论基础和研究现状进行了分析,提出应用领域本体方法建立中医概念信息模型的思路和构想,认为注重研究中医药信息运动规律,包括人体自身信息产生与变化的规律,由自然界信息和社会信息产生对人体影响及其变化的规律,以及人体外在信息、自然界信息和社会信息三者相互影响产生和关联变化的规律等,建立包括这三者内容的中医概念信息模型。

关键词: 中医药 信息模型 领域本体方法

21 世纪以信息科学等为前沿的世界科学技术迅猛发展,自然科学与人文科学间的交叉、渗透、融合,已成为科学技术整体化、综合化发展的重要趋势,信息学方法与技术在中医药学科领域得到广泛的应用和发展,但是由于缺乏中医概念信息模型的研究与突破,顶层设计的理论指导不足,中医药信息研究仍以需求为导向,自下而上的建设,中医药

的特点和优势难以体现,信息离散无序的状态难以得到彻底地改变。信息模型是用以定义信息常规表示方式的方法,建模的对象描述了系统中不同实体、实体行为,及实体间(系统间)数据流动的方式,能够帮助更好地理解系统,用面向对象的方法刻画信息需求,尝试从信息的角度对中医概念进行描述,并从信息的角度建立起中医概念间的关联关系,可以消除中医概念的不确定性,以及中医概念间关联关系的不明确性,是形成中医药信息系统集

收稿日期: 2009-07-24

修回日期: 2009-08-09

* 国家自然科学基金面上项目(30973716):基于领域本体方法构建中医概念信息模型的研究,负责人:谢琪。

** 联系人: 谢琪,博士,主要研究方向:中医药信息学,Tel:010-64014411-3223, E-mail: xieq@mail.cintcm.ac.cn; 崔蒙,本刊编委,博士生导师,主要研究方向:中医药信息学的学科建设,中医药信息数据库与网络建设,中医药信息数据的挖掘与利用研究,以及中医药软科学战略研究, Tel: 010-64013395, E-mail: cui@mail.cintcm.ac.cn。

成的基础和依据。

一、研究现状

科学学科的发展程度取决于社会对它的需求程度,同样,学科理论体系的完善程度也取决于其对社会需求的满足程度,随着信息技术在中医药学的广泛应用,数据库、数据标准、信息平台等实践研究受到广泛重视,但是理论研究较为薄弱,已经难以满足指导研究实践的需要。在中医学与信息学的理论探讨方面,研究认为,中医学和信息学,两者都不是从具体结构上对事物加以解剖性的分析,而是重视从整体上、动态中去观察和研究事物,从而获得关于事物动态现象的运动规律和整体知识,这既是信息方法的功能准则和整体准则,也是中医学的整体观和藏象论。共同的理论基础,相似的方法学使得两个学科在交叉融合过程中逐渐形成了中医药信息学^[1]。在多年理论与实践的研究的基础上,经多次专家论证,我们给定了中医药信息学的定义:中医药信息学是中医学与信息学交叉融合所产生的、研究中医药信息动态现象运动规律的一门交叉科学,其内涵是:中医药信息学以中医药信息为研究对象,中医药领域信息动态现象运动规律为研究内容,应用信息科学方法,实现中医药信息获取、转化与共享,揭示其实质的内在联系^[2]。在中医药学学科中,应用信息学的方法论与技术,我们认为其有着独特研究对象,即中医药信息,它是认识论信息,而不是本体论信息,其特征是与人体相关,在空间上不进行切割,在时间上延续发展,就人体来说是包括人自身经过长期遗传、以及与社会、自然界融合形成的,在人体体表表现出来的综合变化信息,就方药来说是包括中药自身经过长期遗传、以及与自然融合形成的,通过炮制以及配伍作用人体后表现出来的综合信息。它有着独特的研究内容,即中医药领域信息动态现象运动规律。中医药领域的信息动态现象主要包括反映人体整体功能变化的人体外在信息,与人体整体功能变化相关的自然界信息,包括自然环境如季节、昼夜、气候、区

域等和物质如食物、药物等,以及与人体整体功能变化相关的社会信息,如文化与社会关系。因此,可以提出这样的假说:中医药信息运动规律主要包括人体自身信息产生与变化的规律,由自然界信息和社会信息产生对人体影响及其变化的规律,以及人体外在信息、自然界信息和社会信息三者相互影响产生和关联变化的规律等,应建立包括这三者内容的中医概念信息模型。

国内相关研究有:郭玉峰、刘保延等研究了中医本体在中医学术语标准化中的作用^[3],刘和洋、曹宇峰等研究给出了一个病案本体,以及基于该本体的信息抽取方法^[4],曹宇峰、曹存根等研究给出了一个舌诊本体,以及基于该本体的信息抽取方法^[5],王海涛、曹存根等研究给出基于领域本体的信息抽取方法^[6],周肖彬、曹存根等研究给出了一个西医本体,并且探讨基于该本体的知识获取和分析^[7],曹存根等研究给出了一个中医经络系统的本体^[8],任廷革研究提出中医理论模型,即是中医“数据”转化为“信息”,由“信息”再整合成“知识”的过程,接近典型信息过程模型^[9]。

与中医药信息学模型最相关的信息科学领域看,北京邮电大学的钟义信教授提出了“典型信息过程模型”^[10](图1),其含义是当主体要与对象(即系统)打交道时,主体通过获取信息、传递信息和处理信息来认识系统的初始运动状态,达到认知;然后在此基础上“再生”第二类认识论意义的信息(策略信息);后者指明了如何把系统由初始状态转变到目的状态的控制策略;控制的作用则是执行策略信息,产生控制行业,引导系统达到规定的目的状态。有意义的是,任何“对象”都是某种“系统”,因此典型信息过程(信息科学研究内容)模型也是人类认识和改造世界,以及中医诊治疾病的过程。典型信息过程模型揭示了对于对象的运动状态及其变化方式进行优化的

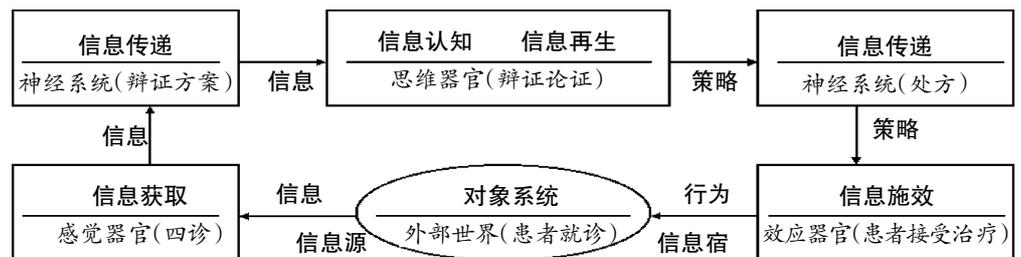


图1 典型信息过程模型

信息过程,即利用信息、通过控制、优化系统,是通用性的模型。

医药卫生领域的模型研究与信息化水平相关,国际上以美国等发达国家现代医学领域的卫生信息模型为主导。国际标准化组织(ISO)健康信息技术委员会(TC215)1994年提出了“卫生信息建模框架”,阐述了“信息模型应当是由概念模型到逻辑模型,最后细化到物理模型”的基本原则,信息建模是一个逐渐细化的过程^[11]。美国 Health Level 7(HL7)是国际知名的卫生信息研究机构,它所研制的参考信息模型(Reference Information Model, RM)被许多国家和跨国企业所参照,或作为元模型所采用以指导信息研究和系统开发。美国 CDC 以 RM 为基础,2002年公布的公共卫生概念数据模型(PHCDM)提出模型的层级观点,将信息模型划分为3个层级^[12],上层是主题域模型,中层是类关系模型、概念数据模型,下层是逻辑数据模型、数据库设计模型和物理数据库模型,这种层级关系与 ISO 提出的层级结构一致,较 ISO 层级更为细化。PHCDM 是多个物理数据库模型的基础模型,它的设计受事物规则和实现模型的实体项目范围所约束,多个下层模型虽然在技术层面不同,如类标识的选择、标准化的程度、关系约束等,但是语义范畴相同。加拿大卫生信息研究院在卫生信息模型研究与发展过程中提出自己的信息模型层级理论,认为信息模型在不同的需求层次将形成不同的模型,依据需求详细程度由粗到细,可划分为:背景模型(领域)、概念模型(一般性理解)、逻辑模型(具体准则)、物理模型(数据存储)^[13]。澳大利亚对模型的层级有与加拿大不同的观点,依据需求详细程度由粗到细,可划分为:概念模型(粗实体)、背景模型(具体属性和关系)、物理模型(数据库设计)^[14]。从各国医药卫生类模型比较,可以看出,模型应根据需求进行分层,由上而下,由粗到细,由抽象到具体在国际已达成共识,不同层级模型可能交互并相互影响,顶层模型较为稳定,底层模型较常调整,但模型分层与名称方面,各国之间有着不同的差异。如 HL7 的 RM 提出主题域有实体(entity)、角色(role),角色链接(role link)、参与(participation)、动作(act)和动作关系(act relationship)6个部分,“动作”主题域是整个模型的核心。加拿大较侧重管理和资源,将其作为重要的主题域单独提取出来。而澳大利亚专家认为,实体间

的关系在较为细节的层次才能够较好地描述,模型的宏观层次应定义为参与者状态、状态变化事件、环境因素、分类体系,而淡化“关系”(联系)。

国际卫生信息模型对国内卫生信息框架产生了较大的影响。2003年卫生部主持的医院基本数据集标准、公共卫生信息系统基本数据集标准体系、国家卫生信息标准基础框架研究,确定了卫生统计指标体系的概念框架的3个维度——健康状况、卫生系统、环境与行为因素。健康状况维度包括人口特征、发病与患病、失能与残障、营养与发育、死亡5类,卫生系统维度包括疾病预防与控制、医疗服务、妇幼保健、卫生监测、卫生资源5类,环境与行为因素维度包括地理与环境因素、社会经济因素、医疗保障、危险因素,并用统一建模语言(Unified Modeling Language, UML)展示不同维度之间、内部指标分类及类间关系^[14]。可以看出,国内目前的模型研制,是从国家统计信息的角度出发,根据卫生统计信息的需求,提出整体框架规划,支持卫生管理和卫生服务决策,不涉及粗粒度的主题域模型,以及中医药信息的研究对象与研究规律。

二、研究设想

用面向对象的方法刻画中医基础理论对概念信息的需求,并同时确定中医基础理论的关键概念信息,明确其主要内容,以形成中医基础理论信息系统集成的依据。研究内容主要包括以下两个方面:

1. 中医药概念信息模型的理论基础研究,支持模型设计与构建

中医药在《黄帝内经》时期,基本理论体系已经形成,两千多年来的发展,这一理论体系在不断的丰富与完善,中医的学科体系是以人为中心建立起来的,以西医以病为中心区别,中医的理论体系不是防病与抗病的理论体系,而是健康的保障体系,人是由精气神组成,精是运动的物质,气是运动的功能,而神“变化出入谓之神”,因此,运动的外在表现可以理解为神。可以看出,中医学通过望、闻、问、切的观察的方法,从整体上,认识并把握到人体环境的关键环节,即运动的状态与方式的过程和规律,从而测知人体内在“精、气”的变化,也因中医的特点是整体的,自然的变化必然影响人体的变化,即“天人合一”。同时中医也是特异性的,由于每个人的个体化的,因此表现在外的运动和方式都是不同的。而中医以

“人”为核心,离不开家庭与社会,因此,对于中医药概念信息模型的理论基础研究,必须联系人文、社会、哲学、自然科学,把握住“人”这个主体及其运动与变化^[15]。

2. 中医药概念信息模型的结构与组成

哲学界的“本体”概念和计算机界的“本体”概念是有区别的。为明确区分二者,哲学领域的“本体”在英文中以大写“O”开头的“Ontology”表示,是一种理论,一种关于存在及其本质规律的系统化解释,这个解释不依赖于任何特定的语言。计算机界的“本体”在英文中以小写“o”开始的“ontology”表示,是一个实体,是对某领域应用本体论方法分析、建模的结果,即把现实世界中的某个领域抽象为一组概念及概念之间的关系的规范化描述,勾画出这一领域的基本知识体系,为领域知识的描述提供含义明确的术语。目前被大部分人公认的定义是 Gruber 在 1994 年提出的:“本体是关于共享概念的一致约定。共享概念包括用来对领域知识进行建模的概念框架、需要互操作的主体之间用于交互的与内容相关的协议,以及用于表示特定领域的理论的共同约定。在知识共享的情况下,本体的形式特化为具有代表性的词汇的定义。一种最简单的形式是一种层次结构,用来详细描述类和它们之间的包含关系。领域知识包括领域概念、概念的性质,概念之间的关系、概念之间的一般规律等,领域本体的研究要求我们根据概念之间的类属关系显式地建立概念之间的联系,明确定义概念所具有的属性、属性的取值约束、处理过程、概念之间的关系等。领域本体还要求明确定义出概念内部或者概念之间的公理,以表示领域内的一般假设或者规律。领域本体的研究使得在人看来一目了然的概念和概念之间的关系都形式化地加以描述,使概念之间的各种规律、联系和假设等都被显式地描述出来,这有利于全面地获取和分析并利用知识。OWL 是 W3C 推荐的一个本体规范。它建立在 DAML+OIL 层之上,其目的在于提供一种语言以描述嵌在 Web 文档和应用中的类及这些类之间的关系(即 Ontology)。OWL 相对于 XML 和 RDFS 拥有更多的机制来表达语义,从而超越了 XML 和 RDFS 仅仅能够表达网上机器可读的文档内容的能力。在信息模型表示方面,OWL(Web Ontology Language)可以作为信息模

型的表示方法。

应用领域本体的知识工程学的方法,构建人、社会、自然界间的中医概念信息模型,将需要解决一系列的关键问题,有如:指导中医药概念信息模型构建的原则是什么?中医药概念信息模型的核心组成是什么?中医药概念信息模型的结构是什么?如何表示这种结构?中医药概念信息模型与西医信息模型,如 HL7 等的关系和差异是什么?我们相信,通过信息建模研究,可以解释中医概念间的关联关系,能够促进中医药信息学理论体系的发展,为指导中医药行业整合离散的科学数据资源,构建面向全社会的网络化、智能化的管理与共享服务体系,实现对中医药科学数据资源的规范化管理及其高效利用提供重要的参考。

参考文献

- 1 崔蒙,尹爱宁,李海燕,等.论建立中医药信息学.中医杂志,49(3):267-269,278.
- 2 崔蒙,尹仁芳,谢琪.关于中医药信息学内涵的研究.中医杂志,2009(10)(稿件录用通知 93794-5).
- 3 郭玉峰,刘保延,李平,等.知识本体与中医临床术语规范化工作.第三届国际中医药工程学术会议论文集,2007,25(7):1368-1370.
- 4 刘和洋,曹宇峰,秦丽娜.基于本体的中医专家临床病案的知识获取方法.计算机系统应用,2005(8):80-83.
- 5 曹宇峰,曹存根.基于本体的中医舌诊知识的获取.计算机应用研究,2006(3):31-34.
- 6 王海涛,曹存根,高颖.基于领域本体的半结构化文本知识自动获取方法的设计和实现.计算机学报,2005,28(12):2010-2018.
- 7 周肖彬,曹存根.基于本体的医学知识获取.计算机科学,2003(1):301-310.
- 8 Cungen Cao, Yuefei Sui. Building an Ontology and Knowledge Base of the Human Meridian-Collateral System. In: Proceedings of the 25th SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence. 2005:195-208.
- 9 任廷革.中医理论信息模型的研究.中国中医药信息杂志,2000,7(11):82.
- 10 钟义信.信息科学原理(第3版)北京:北京邮电大学出版社,2002.
- 11 Australian Institute of Health and Welfare. National Health Information Mode. Canberra: AIHW, 2002:22.
- 12 US CDC. Public Health Conceptual Informational Model (Premiere Edition). Atlanta: CDC, 2002:20.
- 13 Canadian Institute of Health Information Conceptual Health Data Model v2. 3, Ottawa: CIHI, 2001:5.
- 14 王霞,刘丹红,饶克勤,等.国家卫生指标体系:概念框架与基本属性.中国数字医学,2006,1(1):38-41.
- 15 刘保延.用中医发展观看中医发展之路.北京:演讲,2009.1.

Establishment of the TCM Meta Conceptual Model Based on Domain Ontology

Xie Qi¹, Cui Meng¹, Cao Cungen², Liu Baoyan³, Liu Changlin⁴, Jiao Hongguan¹

(1. Institute of Information on Traditional Chinese Medicine, China Academy of Chinese Medical Sciences, Beijing 100700, China;

2. Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China;

3. China Academy of Chinese Medical Sciences, Beijing 100700, China;

4. Institute of Philosophy, Chinese Academy of Social Science, Beijing 100732, China)

Abstract: Informatics methods and techniques have been increasingly developed and applied in the field of the traditional Chinese medicine (TCM). However, due to the lack of research on the conceptual information model of the TCM and the absence of theoretical guidance to the top-layered design, the current TCM information research is usually demand-oriented and can hardly reflect the characteristics and advantages of the TCM. This paper analyzes the theoretical foundation and research status of the TCM conceptual information model, and makes a proposal to establish the conceptual information model of the TCM based on domain ontology. This proposal emphasizes the motion rule of the TCM information, including the generation and variation rule of human self information, the influence and variation rule of human bodies induced by natural information and social information, and the mutual influence and association rule of external information, natural information and social information, based on which the TCM conceptual information model can be established.

Keywords: Traditional Chinese Medicine (TCM); information model; domain ontology method

(责任编辑:李沙沙,责任译审:张立崑)

吉林省科技厅深入长春市调研医药产业发展情况

为切实落实尽快谋划 2010 年及“十二五”重点工作的要求和部署,全面了解长春市医药产业发展现状、发展重点以及相关项目安排情况,7 月 20 日~24 日,由吉林省科技厅副厅长陈维友带队,省科技厅中药办、社发处、专利管理处组成的调研组就医药产业发展情况到长春市进行了调研。调研组先后深入到长春高新技术开发区和经济技术开发区的医药生产企业、大专院校和科研单位进行了实地考察。

在长春高新区,调研组参加了高新区组织召开的专题座谈会,祁健生物、长生生物等 10 余户区内医药企业主要负责同志参加了会议。调研组还到长春长生药业、长春祁健生物和华普药业等企业进行了实地考察;在长春经开区,调研组听取了开发区管委会关于医药产业发展情况的汇报,深入到长春银诺克药业、吉林亚泰生物药业、长春西点药业等企业进行了调研;调研组还到长春生物制品所、东北师范大学、吉林省药物研究院、

吉林农业大学和解放军军事医学科学院十一所等大专院校和科研单位进行了实地考察,详细了解这些单位的医药科研情况。

今年上半年,吉林省医药产业在推进公共创新服务平台、融资服务平台、信息服务平台等方面工作已有效开展。省医药产业发展推进组以吉林省中医药科学院为主体组建了医药公共研发平台;组建了吉林省医药产业发展投资服务有限公司,搭建了省医药产业融资服务平台;加强了医药产业信息服务网络建设等。今后一个时期,省科技厅将突出抓好医药企业的知识产权保护工作,择优选派专利特派员入驻 100 户省内重点医药企业,与企业结成对子,为企业的知识产权保护工作提供全方位服务。近期,还将邀请国家知识产权局的有关领导来长春开展相关培训。

(文 摘)