

# 孕产智能问诊本体知识库研究与应用<sup>\*</sup>

丁腊春 刘朝鸣

(镇江市妇幼保健院 镇江 212001)

〔摘要〕 介绍国内外本体知识库研究现状, 阐述孕产智能问诊本体知识库的构建方法、构建工具及其具体实现, 指出该知识库能为孕产妇提供个性化的诊疗服务, 进一步推动“互联网+医疗”在妇幼保健领域的应用。

〔关键词〕 孕产本体; 知识库; 智能问诊

〔中图分类号〕 R-056 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2018.01.012

**Study and Application of Ontology Knowledge Base of Maternal Intelligent Interrogation** DING La-chun, LIU Chao-ming, Zhenjiang Maternal and Child Health Hospital, Zhenjiang 212001, China

〔Abstract〕 The paper introduces domestic and overseas current situation of ontology knowledge base, expounds on the building method, instrument and specific realization of ontology knowledge base of maternal intelligent interrogation, points out that the knowledge base is able to provide pregnant women and puerperants with customized diagnosis and treatment service and thus further promotes the application of "Internet + Medical" in the field of maternal and child healthcare.

〔Keywords〕 Maternal ontology; Knowledge base; Intelligent interrogation

## 1 引言

随着生活节奏的加快, 休息睡眠的减少, 环境条件的变化, 人们的生活压力和工作压力不断加大, 女性吸烟饮酒也时有发生, 偏食、挑食、营养不良或营养过剩等现象普遍存在, 以至高血压、糖尿病等各种慢性疾病十分普遍, 适龄女性怀孕面临不少困难。同时随着国家二孩政策的全面放开, 不

同体征的妇女都可能有生育第 2 个孩子的计划, 因此高龄高危孕妇比例急剧增加, 剖宫率、早产率等高危因素比例也在不断攀升。孕产领域医疗保健工作已成为妇幼保健工作的重中之重。为此, 亟需找到一种科学、及时、有效的管理方法与管理系统应对这些挑战<sup>[1]</sup>。医疗信息化发展快速, 人工智能 (Artificial Intelligence, AI) 技术层出不穷, 国外已对部分癌症诊疗形成 AI 诊断技术, 如“沃森” (Watson) 机器人可以用自然语言回答专业提问, 对于 7 个癌症病种的辅助诊断已经达到专业水准, 作为一些医院在癌症诊断中的主要参考依据<sup>[2]</sup>。国内也开始在中医领域试点应用, 取得一些成果, 其中“百度医疗大脑”借助云计算、大数据及人工智能技术在传统医疗领域进行尝试应用<sup>[3]</sup>。然而孕产管理领域仍停留在独立的医疗系统内, 形成较为严重的信息孤岛, 大量的诊疗数据没有被有效地整

〔修回日期〕 2017-11-15

〔作者简介〕 丁腊春, 高级工程师; 通讯作者: 刘朝鸣, 馆员。

〔基金项目〕 2016 年镇江市重点研发计划 (项目编号: SH2016004); 2017 年江苏省重点研发计划 (项目编号: BE2017700)。

理、分析与利用。因此镇江市妇幼保健院在融合诊疗指南及文献知识的基础上尝试建设孕产本体知识库,以应对这些挑战。

## 2 国内外本体知识库研究现状

### 2.1 国外

国外对本体的研究非常活跃,20 世纪 60 年代末期自 Collins、Quillian、Loftus 等人开始研究,20 世纪 70 年代初 Simon、Schank、Minsky 等人陆续提出一些相关理论并取得一定成果。其中 Simon 在进行自然语言理解的应用研究时提出语义网络 (Semantic Network, 不同于 Semantic Web) 概念,2001 年在互联网创始人 Tim Berners Lee 的倡导下,W3C 联盟建立语义网行动计划<sup>[4]</sup>,其推荐的网络本体语言 (Web Ontology Language, OWL) 标准已成为本体领域知识表示标准,成为各种本体知识库共享与复用的标准工具<sup>[5]</sup>。通过进一步规范与推广应用,现已形成较为成熟、可靠、易检索的知识库建设平台。

### 2.2 国内

国内对于本体的研究与国外相比起步较晚,从最初的翻译国外文献,到各医学领域的专题研究,已经取得不少研究成果,特别是近几年智慧医疗建设发展迅速,医疗领域本体应用研究也越来越多,如 1995 年曹存根教授还用本体理论提出一项新课题——国家知识基础设施 (National knowledge Infrastructure, NKI),其中 NKIMed 是 NKI 的一个子集<sup>[6]</sup>;2001 年国家科技部支持浙江大学计算机学院自主研发语义网格软件系统,主要围绕中医药信息化和智能交通等应用领域<sup>[7]</sup>,形成一套语义网格理论体系,实现语义网格在中医药知识共享与服务中的应用<sup>[8]</sup>。

由于这些知识库没有专门针对孕产妇医疗保健的内容,也没有提供统一的网络访问系统,故本文提出以移动互联网、人工智能、大数据挖掘等技术为支撑,以孕产妇孕前、孕中、产后为时间轴,融合医院海量文献、诊疗指南和个人健康档案数据,为孕产妇

提供智能问诊服务。其关键核心内容之一是设计并建立一种孕产智能问诊知识库。

## 3 孕产智能问诊本体知识库构建过程

### 3.1 方法

本体对于知识库十分重要,它是一个领域知识库系统的基础。构建一个孕产领域知识库应用本体,首先理清所需要的词汇、层次结构以及词汇之间的关系,其次构建可以被共享和利用的孕产知识库框架。孕产知识库的本体框架至关重要,直接影响到知识库共享和复用体验效果。不同方案构造本体的过程也会各不相同,分析研究目前较常用的本体构建方法,包括 7 步法、骨架法和工程法等。在医疗领域常采用的构建方法是由斯坦福大学提出的 7 步构建法,简称 7 步法;在企业领域常采用的是由英国爱丁堡大学推出的骨架构建方法,称之为骨架法;在智能图书馆领域常采用支持 EXPRESS 和 Ontolingua 两种语言的工程法。经过对比采用 7 步法构建孕产智能本体。

### 3.2 工具

利用 Protégé 工具导入孕产实例构建知识库,构建与孕产相关的不同知识库,如病例库、疾病库、症状库、指南库、文献库、检查检验库、药物知识库等,最终共同形成孕产智能问诊的知识库。知识库中知识单元之间存在一定的关联关系,1 种疾病可以按照不同角色、不同阶段、不同治疗手段划分为多个知识单元,疾病的各个检验检查指标、症状、诊疗指南及文献、病案之间存在联系,各种治疗手段之间也因治疗疾病而存在联系。患者的运动、睡眠、营养数据等作为数据属性,与其他对象不发生直接联系,只是一种佐证材料;检验检查报告数据、病理报告数据等作为关联数据,与疾病的诊断、治疗方案存在关联关系,是知识库的核心。疾病库中的临床表现与症状库建立相互关联关系,通过链接疾病库的检查检验报告指南即可以得到检查检验指标知识库,通过与检查检验指标的关联比对,知识库会推荐与检查检验指标相对应的疾病,再从疾病

知识库中获取该疾病的检查检验指标知识库，彼此验证结果。各知识库之间底层关联，见图 1。

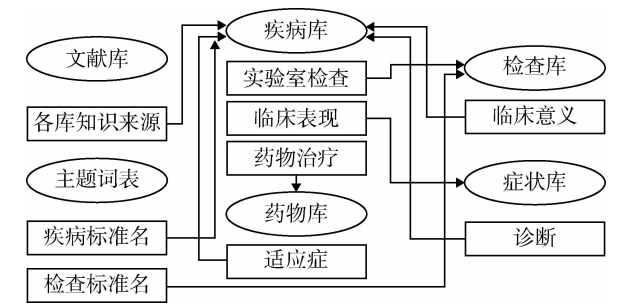


图 1 知识库之间底层的关联实现

3.3 孕产本体构建

3.3.1 总体架构（图 2）

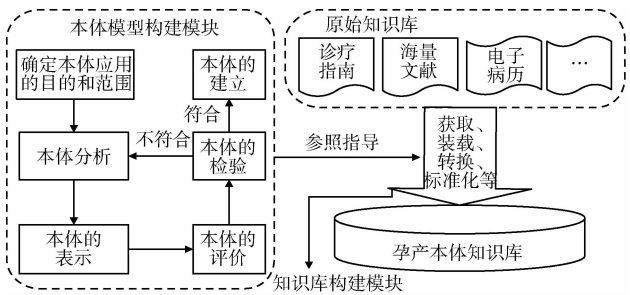


图 2 孕产本体知识库模型总体架构

孕产智能问诊知识库能够为孕产妇提供个性化

诊疗服务，孕产妇在围产期时可能会有某些相同的病症，如妊娠性糖尿病、妊娠性高血压等，常伴有孕吐、乏力、羊水异常等临床症状，因此可以从其症状表现、辅助诊断方法、运动健康指导、用药指导、康复保健指导等方面来进行本体构建。7 步法可实现对围产期全生命周期的保健、治疗、康复等关键指导并进行疾病的分析与探讨。

3.3.2 7 步法构建孕产本体流程 第 1 步明确构建领域及其覆盖范围。本研究以孕产领域为构建领域，镇江市妇幼保健院是市三级甲等妇幼保健医院，具备相关文献、指南等知识平台和孕产妇个人健康档案平台。通过知识平台和个人健康档案平台，为本体知识库的构建提供坚实数据支撑基础。第 2 步通过对领域内医疗知识本体的检索与研究，未发现有可借鉴复用的类似孕产本体知识库。第 3 步提炼出与孕产智能问诊相关的关键概念术语。通过对文献、指南的学习与研究，构建孕产本体的顶层概念：孕产疾病类型、症状与体征、检验检查、疾病用药等，整体数据元之间的关系网络，见图 3<sup>[9]</sup>。第 4 步定义类及类的等级关系。第 5 步定义类及其等级分类的属性。第 6 步定义属性的分面。第 7 步建立实例。通过对孕产领域文献及诊疗指南等知识平台的研究，以及孕产妇自身的个人健康档案平台来提取相关的孕产领域实体，完善孕产智能问诊领域本体实例<sup>[10]</sup>。

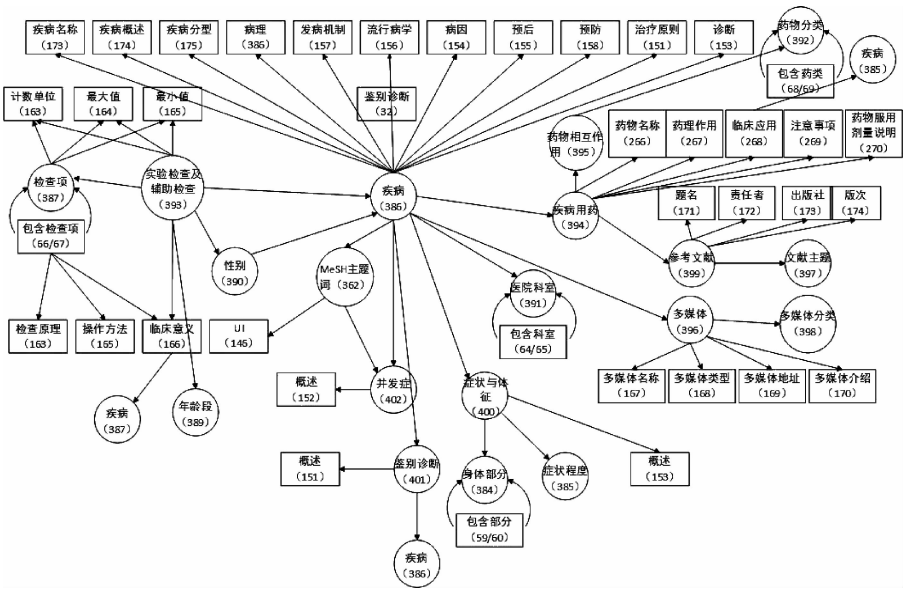


图 3 知识库数据元之间的关系网络

3.3.3 数据抽取 孕产本体知识库数据首先从诊疗指南、文献指南、检验检查指南、药理指南的文本中抽取关联的事件、事实等信息，即进行分词分析，将这些分词进行结构化处理。其次从各类围产期健康保健书籍中抽取孕产妇的健康体征指标、康复保健方法等信息分词进行结构化处理。再次从临床病案中抽取典型病历的症状、治疗方法、康复保健指导等临床路径信息进行结构化处理。最后按照抽取诊疗指南、文献、检验检查指南、药理指南、典型病案中的结构化、非结构化文本进行分词结构化并统一归档，初步形成孕产本体库。数据抽取的流程，见图 4。

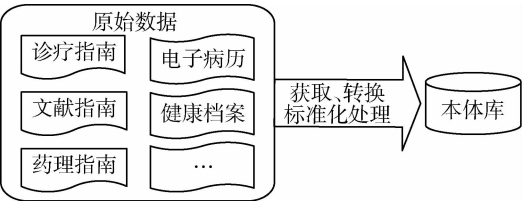


图 4 数据获取流程

以海量文献为例，各类文献以公开文本资料记

载，其中患者临床表现、医生诊断、临床护理记录、康复记录等多项患者统计信息汇总在一起，每位作者根据规律提出不同的观念与结论，供他人分享参照。而本体的复用是要将多篇文献的类似规律找出来，整理出临床路径原型，再根据临床路径原型创建孕产文献库复用框架。

3.3.4 本体知识库建立 以孕产学科为主线，分别建立与之相关的病例库、疾病库、症状库、指南库、文献库、检查检验库、药物知识库等不同的知识库，通过本体技术，共同形成孕产学科的知识库，为智能问诊提供数据知识支撑。

3.3.5 智能问诊服务 构建孕产知识库模型，形成专科特征的知识平台，为孕产妇提供个性化的诊疗服务。孕产妇无需到医院，可利用网络连接到智能问诊系统中，智能问诊平台通过患者主索引调阅该孕妇的所有历史信息到平台实时处理模块等待查阅。平台通过对孕产妇自然语言进行后结构化处理、语义识别，再结合个人健康档案、历史咨询问题对孕产知识库库进行智能搜索筛选，为孕产妇优选推荐个性化诊疗方案。智能问诊架构，见图 5。

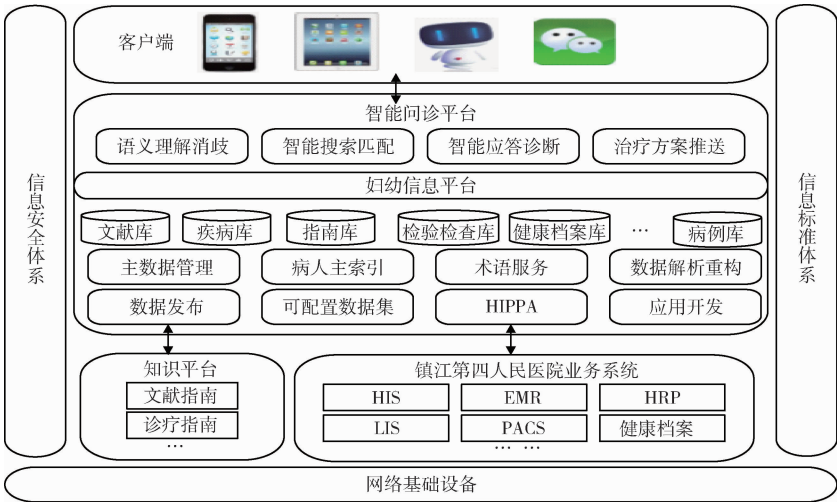


图 5 孕产智能问诊整体架构

4 结语

在镇江市妇幼保健院上线的孕产智能问诊知识库，整合海量文献、诊疗指南构建孕产领域知识库模型，初步为 4 类妊娠状态的 3 类孕产人群提供知

识服务。目前已有 5 000 名注册孕产妇，112 名注册医生，问诊及咨询量达 3 000 人次，点击量 100 万，使用效果达到预期设想。后期将在知识库的基础上逐步实现为孕产妇提供结合个人特点及历史状

(下转第 76 页)

## 参考文献

- 1 代涛, 钱庆, 王小万, 等. 医疗卫生领域知识服务与知识管理的理论和实践 [J]. 医学信息学杂志, 2008, 29 (4): 1-10.
- 2 张晓林. 走向知识服务 [M]. 成都: 四川大学出版社, 2001: 110-112.
- 3 张红丽, 吴新年. 知识服务及其特征分析 [J]. 图书情报工作, 2010, 54 (3): 23-27.
- 4 刘佳. 高校图书馆学科知识服务模式研究 [D]. 长春: 吉林大学, 2007.
- 5 董玮, 詹庆东. 图书馆知识服务模式辨析 [J]. 图书馆学研究, 2016, (3): 72-79.
- 6 张海涛, 宋拓, 刘健. 高校图书馆一站式知识服务模式研究 [J]. 情报科学, 2014, 32 (6): 104-108, 113.
- 7 马天舒. 大学图书馆开放型知识服务模式研究 [J]. 情报杂志, 2013, 32 (2): 135-138, 102.
- 8 张晓林. 走向知识服务: 寻找新世纪图书情报工作的生长点 [J]. 中国图书馆学报, 2000, 26 (129): 32-37.
- 9 陈昊天. 跨界搜索对服务创新绩效的影响研究 [D]. 杭州: 浙江理工大学, 2016.
- 10 张美莉. 基于“互联网+”思维的图书馆跨界服务研究 [J]. 情报杂志, 2016, 35 (9): 204-207, 198.

- 11 Atkinson R W, Kang S, Anderson H R, et al. Epidemiological Time Series Studies of PM2.5 and Daily Mortality and Hospital Admissions: a systematic review and meta-analysis [J]. Thorax, 2014, 69 (7): 660-665.
- 12 Franchini M, Guida A, Tufano A, et al. Air Pollution, Vascular Disease and Thrombosis: Linking Clinical Data and Pathogenic Mechanisms [J]. Journal of Thrombosis and Haemostasis, 2012, 10 (12): 2438-2451.
- 13 Shang Y, Sun Z, Cao J, et al. Systematic Review of Chinese Studies of Short-term Exposure to Air Pollution and Daily Mortality [J]. Environment International, 2013, 54 (4): 100-111.
- 14 屈芳. 环境气象因素对呼吸系统疾病影响的研究进展 [J]. 气象科技进展, 2013, 3 (6): 35-44.
- 15 赵桂柱, 张冬梅. 论肿瘤与环境的关系 [J]. 中国中医药现代远程教育, 2012, 10 (12): 91-92.
- 16 郭堂春. 华中区域常见慢性非传染性疾病前瞻性队列研究 [J]. 中华预防医学杂志, 2016, 50 (10): 927.
- 17 美国疾病控制与预防中心 [EB/OL]. [2017-04-20]. <https://www.cdc.gov/>.
- 18 陈松景, 吴思竹, 侯丽. 基于疾病与环境科学数据的跨领域关联分析及应用 [J]. 中华医学图书情报杂志, 2017, 26 (7): 1-6.

(上接第 54 页)

态的健康建议方案, 推广“专科特征+平台互动”的健康互联网应用, 推动“互联网+医疗”在妇幼保健领域的应用, 为智能问诊软件系统开发和孕产智能陪护机器人实体研发奠定核心基础。由于孕产智能问诊知识库是从科研成果转化到实践应用, 需要一个逐步完善与应用推广的过程, 需经历较长时间的不断知识积累与实践验证才能形成可靠有效的应用。

## 参考文献

- 1 丁腊春, 王译, 石雷, 等. 面向孕产妇的移动医疗健康平台“孕乐宝”开发与应用 [J]. 医学信息学杂志, 2016, 37 (10): 40-43.
- 2 吴佳男. 智能医疗: 或成医疗机构新突破口 [J]. 中国医院院长, 2017, (2): 89.

- 3 常晓. AI+医疗 [J]. 人人健康, 2016, (21): 22-23.
- 4 瞿裕忠, 胡伟, 程龚. 语义网技术体系 [M]. 北京: 科学出版社, 2015: 10.
- 5 邱旭. 领域本体在糖尿病领域知识服务中的应用 [D]. 哈尔滨: 哈尔滨工业大学, 2013.
- 6 李博. 基于语义关系的高血压临床指南知识库构建 [D]. 成都: 电子科技大学, 2013.
- 7 陈华钧, 姜晓红, 吴朝晖. 支持中医药信息化的语义网格平台实现 [M]. 杭州: 浙江大学出版社, 2011: 6.
- 8 吴朝晖, 陈华钧. 语义网格: 模型、方法与应用 [M]. 杭州: 浙江大学出版社, 2008: 5.
- 9 范启勇. 上海市医院信息集成平台建设与实践应用指南 [M]. 北京: 科学出版社, 2016: 11.
- 10 来建梅. 基于领域本体的中医骨伤知识库的研究与构建 [D]. 济南: 山东中医药大学, 2013.