

基于本体的孕产知识库构建及应用*

王 译 丁腊春 叶恩光

(镇江市第四人民医院 镇江 212001)

[摘要] 将本体、知识库与孕产疾病知识结合构建基于本体的孕产知识库, 阐述具体构建步骤, 包括明确本体领域范围, 列举概念术语, 定义类、属性及类之间的关系, 创建本体实例等方面, 促进孕产知识复用与共享。

[关键词] 孕产妇; 本体; 孕产知识库; Protégé; 网络本体语言

[中图分类号] R-056 **[文献标识码]** A **[DOI]** 10.3969/j.issn.1673-6036.2018.06.012

Building and Application of the Ontology - Based Maternity Knowledge Base WANG Yi, DING La - chun, YE En - guang, The Fourth People's Hospital of Zhenjiang, Zhenjiang 212001, China

[Abstract] To build the ontology - based maternity knowledge base by combining ontology, knowledge base and maternity disease knowledge, the paper elaborates on the concrete building steps, including validating the scope of the ontological domain, enumerating concepts and terms, defining categories, properties and relationship among categories, creating ontology instances and other aspects to promote the multiplexing and sharing of maternity knowledge.

[Keywords] Pregnant women; Ontology; Pregnancy - labor knowledge base; Protégé; Ontology Web Language (OWL)

1 引言

随着国家二胎政策的全面放开, 孕产妇的数量日益渐增^[1]。现有孕产妇信息管理大多仍停留在独立的系统内, 形成较为严重的医疗信息孤岛, 大量的诊疗数据、海量文献等缺乏有效地整理、分析和利用。对其进行数据挖掘和知识发现, 实现孕产领域信息共享, 是提高孕产保健工作水平的关键^[2]。

国内对本体的研究已有一定的成果, 但在孕产领域尚未有规模化、可共享、可复用的医学知识库^[3], 因此本研究在已有的海量文献、指南等数据基础上, 利用本体技术和图形化建模工具, 实现孕产知识库构建, 为临床决策支持和智能问诊提供数据知识支撑。

2 孕产知识库模型架构

孕产知识库以孕产妇孕前、孕中、产后为时间轴, 融合医院海量病历信息、文献和诊疗指南, 集合围产期保健各种疾病症状、诊断、治疗等知识, 构建孕产主要疾病的知识系统, 通过语义识别、智能搜索等技术实现智能人机交互, 为临床医生和孕产妇提供知识支持, 孕产知识库模型框架, 见图1。

[收稿日期] 2018-02-08

[作者简介] 王译, 硕士, 工程师; 通讯作者: 丁腊春, 高级工程师。

[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)。

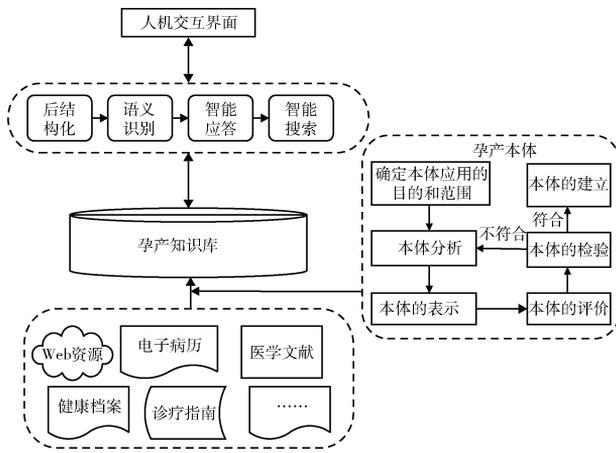


图1 孕产知识库模型框架

3 基于本体的孕产知识库构建

3.1 概述

构建一个可复用、可共享知识库的关键是要设计出相应的本体。目前已有不少本体构建的方法，如骨架法、7步法、评价法。构建工具有 Protégé、OntoEdit、Apollo 等^[4]，经过优缺点分析，本研究借鉴7步法的方法和流程，结合自身学科特点，使用 Protégé 构建工具和网络本体语言（Web Ontology Language, OWL）进行孕产本体构建^[5]。

3.2 明确本体领域和范围

随着二胎政策的全面放开，涌现大批高龄孕产妇，各种妊娠合并症及并发症发生率不断增加，因此大量学者致力于研究孕产保健领域并且著有大量的文献和书籍，如由人民卫生出版社出版的《妇产科学（临床版）》^[6]。通过学习这些文献，结合妇产科医生的大量临床经验，了解孕产领域相关知识，如孕产疾病的病症表现、疾病病因、症状之间的因果关系等，为后续孕产本体的构建奠定知识基础。

3.3 列举概念术语

创建孕产本体之前首先要确定本体领域中的关键概念，对孕产概念进行标准化，以便后期建立的孕产知识库能够复用和共享。本研究通过对孕产领域专家的调查研究以及文献的学习发现，孕产疾病都具有相同的基本特征，都可以从疾病、病因、症

状、临床检验检查、治疗、保健等方面来描述。因此通过对这些通用特征进行分析，列举本体中的相关术语：（1）疾病。孕产妇在围产期保健期间常见的疾病有170种左右，如妊高症、妊娠合并糖尿病、脐带脱垂、早产等。（2）病因。遗传、内分泌、饮食、子宫胎盘因素等。（3）症状。不同的孕产疾病出现的症状不同，如妊高症有水肿、头痛、眼花等症状。妊娠合并糖尿病有多食、多饮、多尿、体重减轻、呕吐等症状。（4）临床检验检查。一般有抽血、验尿、B超等。（5）治疗。西医治疗、膳食方案、终止妊娠。（6）保健。定期孕期检查、加强孕期保健、运动、睡眠及营养管理。

3.4 定义类、属性以及类之间的关系

3.4.1 构建类 本研究采用自顶向下的方法构建本体类的层次结构，在孕产本体中分别建立疾病种类（Diseases）、病因（Cause of disease）、症状（Symptom）、临床检验检查（Clinical examination）、治疗（Treatment）、保健（Health-care）6大类。在类的构建过程中，有的类之间存在互斥关系，如病因和治疗这两类是不会存在相同的实例，导致疾病的因素永远不会是治疗疾病的某种方法，所以还需建立相应的互斥关系，病因和治疗类的互斥关系，见图2。

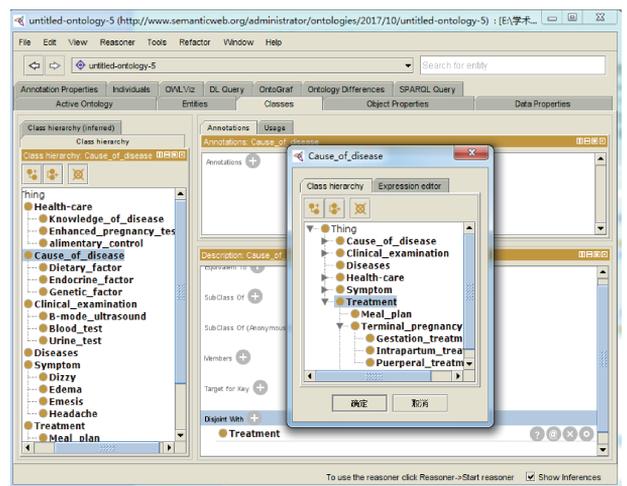


图2 病因和治疗类的互斥关系

3.4.2 定义属性 通过定义属性来确定孕产本体各个类之间的关系。孕产本体模型，见图3。提供计算机可以理解的语义信息。本体之间的关系主要

有 is - a, instance - of, attribute - of。本研究中描述类特征属性的主要有 is - a、caused - by 与 causes 关系, has - symptom 与 is - symptom - of 关系, has - treatment - 与 is - treatment - of 关系等, 对它们的逆关系、定义域、值域、互斥性等进行描述, 最后利用自带的推理机对本体概念的层次结构、一致性进行检验推理, 从而完成孕产本体模型的构建。

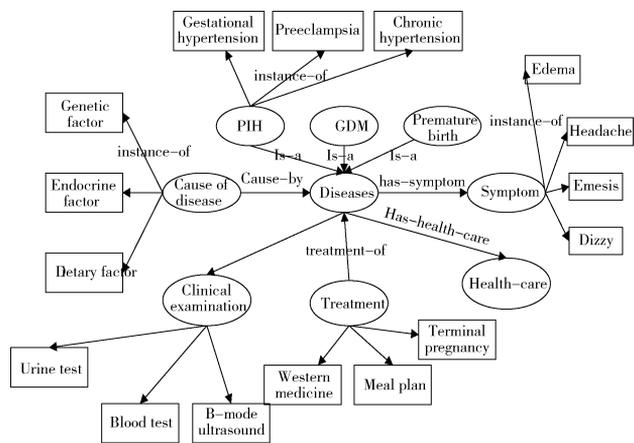


图 3 孕产本体模型

3.4 创建本体实例

建立孕产本体模型后, 可在本体定义的类中添加相应的实例, 包括实例名称、属性和属性值等。当大量附有属性和属性值的实例被添加后, 导入数据库中就形成相关的知识库。同时针对一些实例以结构的形式展示出来, 由于关联的实例数较多, 仅列举一部分^[7]。

```

疾病类型: 妊高症
{
has - symptom: 水肿; 尿蛋白; 不正常的血收缩压或舒张压……
causes: 子痫前期与子痫、胎盘早剥、产后大出血……
has - treatment: 硫酸镁等综合西医治疗; 中医按摩……
}
症状: 不正常的血压值
{
is - symptom - of: 妊高症
is - clinical - examination - by: 血压测量; 尿常规……
}
    
```

3.5 具体应用

通过 Protégé 构建的孕产本体是以 OWL 文件形式存储, 如需知识更好地应用和共享, 形式化的本体就需要一种与平台无关、独立持久的存储格式, 这样本体才能以明确且统一的语义表达在不同的应用中。本研究选择 MySQL 数据库存储孕产本体, 通过数据库存储方式便于存取与查询。在孕产知识库的基础上设计“小医在线”智能对话程序并集成在医院拥有自主知识产权的“孕乐宝”APP^[1]上, 为医生、孕产妇提供孕产知识查询。如用户输入“妊高症是什么”, 系统对用户的自然语言进行处理, 通过语义识别、智能搜索等技术分析出相关实例, 返回检索结果, 通过关键字“妊高症”推送一些相关的其他问题。

4 结语

本研究构建的孕产知识库较普通数据库具有更多的语义信息, 能够更好地实现孕产知识的复用和共享。因时间、技术等因素, 目前只实现孕产知识库的构建, 且规模较小, 后期将继续建立规则推理, 最终实现智能问诊。

参考文献

- 1 翟振武, 张现苓, 靳永爱. 立即全面放开二胎政策的人口学后果分析 [J]. 人口研究, 2014, 38 (2): 3 - 17.
- 2 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.
- 3 张斌, 魏扣, 郝琦. 国内外知识库研究现状述评与比较 [J]. 图书情报知识, 2016, (3): 15 - 25.
- 4 王向前, 张宝隆, 李慧宗, 等. 本体研究综述 [J]. 情报杂志, 2016, 35 (6): 163 - 170.
- 5 王美琴, 吴庆斌. 基于本体的医学知识库构建方法综述 [J]. 医学信息学杂志, 2017, 38 (3): 73 - 76.
- 6 曹泽毅. 中华妇产科学 (临床版) [M]. 北京: 人民卫生出版社, 2010.
- 7 中华医学会妇产科学分会妊娠期高血压疾病学组. 妊娠高血压期疾病诊治指南 (2015) [J]. 中华妇产科杂志, 2015, 50 (10): 721 - 728.