

基于本体的医学知识库构建方法综述^{*}

王美琴 吴庆斌

(暨南大学附属第一医院信息科 广州 510630)

〔摘要〕 围绕基于本体构建医学知识库流程中的本体构建、诊断规则构建、本体推理机推理实现几方面,分别阐述各个环节中所涉及的关键技术和方法,列举用相关技术构建领域知识库的实例,探索构建医学领域知识库的最佳方法。

〔关键词〕 医学知识库; 本体; 诊断规则; 推理引擎; 综述

〔中图分类号〕 R-056 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2017.03.016

Review on the Construction Methods of Ontology-based Medical Knowledge Base WANG Mei-qin, WU Qing-bin, Department of Information Technology, the First Affiliated Hospital of Jinan University, Guangzhou 510630, China

〔Abstract〕 Around the ontology construction, diagnosis rule construction, ontology inference mechanism and so on in the process of ontology-based construction of medical knowledge base, the paper states respectively the involved key technologies and methods in each link, lists the examples of the construction of the domain knowledge base through relevant technologies, explores the best method of constructing the medical domain knowledge base.

〔Keywords〕 Medical knowledge base; Ontology; Diagnosis rules; Inference engine; Review

1 引言

现代科技的发展,造成了所有领域的知识大爆炸,原有的知识组织体系已无法满足需求,如何有效组织和利用知识成为摆在科技工作者面前的重要课题。在医疗卫生领域,为有效获取、储存、管理及应用医疗机构产生的数据资源,以达到对疾病的管理、控制和医疗研究的目的,医学知识库应运而

生。我国医学知识库研究最早源于 20 世纪 70 年代末的中医专家系统^[1]; 20 世纪 90 年代以来,广泛服务于临床研究的机构知识库是我国医学知识库发展的另一重要阶段; 进入 21 世纪,我国医学知识库研究进入了全新的发展时期; 目前知识库研究已逐渐成为医学信息学领域的关注热点。

在区域医疗与健康大数据中,80% 的数据是以非结构化形式存储的,要想很好地运用并非易事^[2]。通常这些数据都是采用自然语言处理(Natural Language Processing, NLP),因此如何对自然语言进行语义分析(Semantic Analysis, SA)和数据特征提取,选择适当的数据模型进行计算和验证,是用好医疗大数据的关键所在^[3]。目前,基于本体的知识库构建思路和方法已成为全世界知识库研究的主要方向^[4],被广泛应用于语义 Web 智能信息检索、信息集成、数字图书馆等领域。本体的应用可以解决医

〔收稿日期〕 2016-12-19

〔作者简介〕 王美琴,硕士,助理工程师,发表论文 2 篇; 通讯作者: 吴庆斌,高级工程师。

〔基金项目〕 广东省省级科技计划项目(项目编号: 2016B010108008); 广东省重大科技专项(项目编号: 2015B010106008)。

学临床用语的通用性和规范化难题，尤其是中医临床用语本身存在语义模糊和关联复杂等问题^[5]，已成为医学知识库构建的通用工具。

通过对本体及其相关技术的研究，构建知识库的基本模型架构，主要包括本体库、规则库及推理引擎。本体库是根据临床知识抽取出的知识元素所构建的本体；规则库是由临床专家根据临床经验总结出的临床常见规则；推理引擎通过对本体及规则的推理得出新的知识，再进一步对本体进行修正。本文在收集我国医学知识库研究文献的基础上，总结分析医学知识库的构建方法，以期构建最佳医学知识库提供参考。

2 构建医学本体库

2.1 医学本体库构建流程

构建医学本体库的第 1 步是依据医学学科的内容，形成本学科的基本框架；接着，依据建立好的体系框架，提取框架内所有的概念，定义概念与概念之间的关系^[6]；然后，选择合适的本体构建方法或者工具对建立的领域本体进行编码以及形式化，医学本体的检验需要专家的参与，本体库也需要在系统的使用过程中不断地加以完善和维护^[7]。本体库构建的流程，见图 1。

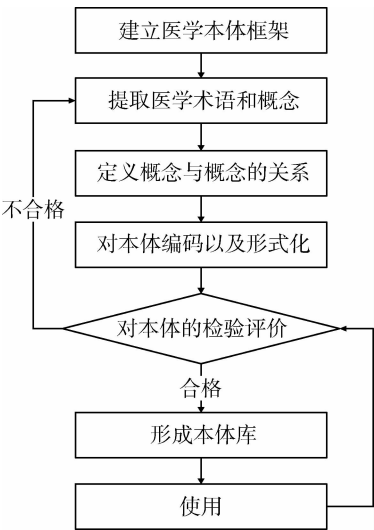


图 1 本体库构建流程

2.2 本体构建方法

目前大部分本体的构建都是基于特定领域，所

以需要相关领域专家参与，构建方法也是针对特定项目提出来的。其中比较著名的有骨架法、七步法、KACTUS 工程法，METHONTOLOGY 方法以及企业建模法。从软件开发和项目管理的角度评价本体的构建方法，METHONTOLOGY、七步法涉及到了生命周期的概念，其中 METHONTOLOGY 更符合软件开发的思想。从应用的领域范围分析，骨架法和 KACTUS 工程法仅适用于 1 个领域，其他几种方法可以应用于多个领域^[8]。七步法对具体的方法细节做了详细的描述，这为开发人员构建本体提供了全面的指导依据。每一种方法都具有各自的特点，开发人员需要根据具体的研究领域选择适合自己的本体构建方法。李博等^[8]选用了骨架法构建高血压临床指南本体，巩沐歌等^[9]将骨架法和七步法两种方法结合起来，充分利用两种方法的优势，以此提出了一种新的构建本体方法。

2.3 本体构建工具

当前，国外许多大学和研究机构正在研究和开发的本体构建工具很多，到目前为止，本体构建工具的总数超过了 90 个，比较成熟、知名度高、较为常用的主要有 Protégé、KAON、WebODE、Ontoedi 等。其中，Protégé 在支持中文方面具有基于 UTF-8 编码、开源、良好的可扩展性优势，所以其使用率较高。如徐彬峰等^[6]使用 Protégé 建立冠心病医学本体知识体系、李新霞^[10]利用 Protégé 构建中医脾胃病本体模型、冯贞贞等^[11]利用 Protégé 本体构建工具进行 5 个病种本体的构建、李博等^[8]利用 Protégé 工具构建高血压临床指南本体、罗雪琼等^[12]使用 Protégé 构建泌尿外科疾病医学知识库、苟玲等^[13]利用 Protégé 构建基于本体的抗生素知识库等。

3 构建医学诊断规则

在用本体构建工具构建好类、属性和实例之后，基本的本体库就已建好，接着就需要构建诊断规则库。规则是从人们的实践经验中提炼总结出来的知识，能指导专家系统由已知事实推导出新事实

或者某种结论。利用语义 Web 规则语言^[14] (Semantic Web Rule Language, SWRL), 在本体中加入规则, 可以提供更强的逻辑表达能力, 构建的本体表达的内容更为丰富, 实用性更强。SWRL 规则通常写成“前提 $\rightarrow$ 结论”的形式, 即 $X_1 \wedge \cdots \wedge X_n \rightarrow Y$ 。

4 本体推理机

本体推理机的一般系统结构由本体解析器、查询解析器、推理引擎、结果输出模块和应用程序接口 (Application Program Interface, API) 5 大模块组成^[15], 见图 2。常用的推理机有 Pellet、RACER、FaCt++、Jess 以及 Jena 等。Jena 推理机^[16]推理过程中所需要的时间比较多, 因此运行效率低是一个很明显的缺陷。该推理机^[17]对本体描述语言的依赖性很强, 只能用于特定的几种本体描述语言实现推理功能, 而用 Jess 的空间效率很低, 推理机制很难优化。每一种推理机针对的领域本体不尽相同, 开发人员需要根据具体的研究领域选择适合自己的推理机。李博等^[8]结合高血压临床指南本体的特点选用了 Jess 推理机, 冯贞贞等^[11]在 Protégé 上用 Jess 推理引擎实现了 5 个病种的临床路径知识的推理, 李新霞等^[10]通过 Jena 技术完成对中医脾胃病本体的解析, 苟玲等^[13]通过 Jena 本体引擎解析抗生素知识库本体文件。

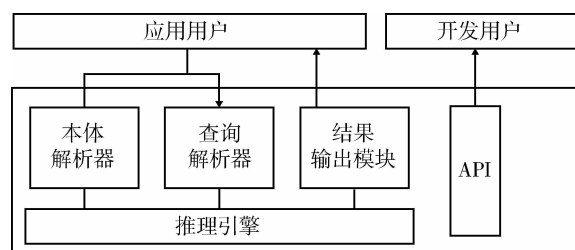


图 2 本体推理机系统结构

5 推理实现

完成了本体和诊断规则的构建后, 接着就要实现诊断规则和本体的推理实现。在整个推理实现过程中, 共分为 4 步进行, 分别为检测冲突、优化表

达、本体合并和实例归类。推理的一个基本内容就是由给定的知识获得隐含的知识, 在本体中的推理从根本上说就是将隐含在显式定义和声明中的知识通过一种处理机制提取出来。以本体推理机实现诊断推理功能, 将患者的临床数据传送到本体库中, 成为本体中的实例, 然后进行推理, 得出推理结果。但有时本体语言不能被推理机所识别, 需要借助相关的解析器或者插件完成映射。如借助于 Protégé 自带的 SWRLTab, SWRLTab 插件可以实现 SWRL 规则到 Jess 规则库的映射^[18]。范轶^[19]用心电图本体提供推理事实的基础, SWRL 提供推理所使用的规则, 通过 Protégé 提供的将 OWL 本体 + SWRL 转换为 Jess 功能按钮, 将捆绑了 SWRL 推理规则的 Electrocardiogram.owl 映射到 Jess 规则库, 形成 Jess 规则供推理机使用。肖敏^[20]使用 SWRL 规则语言, 借助 SWRLTab, 详细地描述基于生理参数的疾病诊断预警规则, 设计并实现诊断预警推理规则库, 使用 Jess 引擎对推理规则进行试验, 保证了 SWRL 规则的正确性和推理的有效性。许德山等^[21]在分析本体在知识检索中的作用时, 也使用了 Protégé 中的 SWRLTab 插件调用外部推理机的高效接口, 因为利用该插件比单纯使用 Jess 引擎更加方便^[22]。

6 展望

6.1 医学术语标准化

在医疗卫生服务活动中, 合理的医疗决策要依靠内容准确、数据一致、结构完整和实时的临床信息, 这些临床医疗信息是经过临床医疗机构的数据采集、归类整理、后期提炼等过程才能获取。由于在发展过程中受诸多因素的影响, 临床症状存在名称繁多、内涵和外延不清、表述与理解不统一的情况, 要想解决症状体征术语不全、映射不统一等问题^[23], 必须修正临床术语的概念表、描述表和关系表, 建立统一规范的医疗信息标准^[24]。

6.2 制定诊断规则

同样的临床问题, 不同国家或同一国家的不同地区甚至在一个地区的不同社区, 其处理方式也大

相径庭。医生对患者的诊断过程主要依据和参考两方面的知识^[8]：一方面是医学书籍和医学期刊的知识，这类知识是医生临床实践过程中的理论基础；另一方面是经验性知识，这类知识主要是医生根据临床经验的积累和归纳得到的。因此，知识库规则的构建是一个反复循环的过程，需要在使用过程中根据相关专家的经验不断总结，进行修改和完善，使其越来越满足实际应用的需求。

7 结语

构建本体知识库不可能一蹴而就，是一个不断完善、反复验证的过程，特定领域本体知识库的初始版本构建后，需要不断的排错，其最大优势在于对领域知识实现动态、开放的组织，达到更高效的信息共享。在构建医学本体知识库的过程中，要充分考虑显性知识（文档、数据、已有数据文献）的深入挖掘、隐性知识（专家经验、临床实践、个人总结）的整理发现。同时，针对目前我国知识库开发处于研究相对分散、功能较为单一的情况，采用何种知识组织和表示方式实现知识共享和二次利用，为临床医生和研究人员提供长期、动态和全面的知识源，需要投入大量且长期的人力和物力去不断探索和完善。

参考文献

- 1 龙远雄, 马利, 王彬, 等. 近 10 余年来国内医学知识库研究文献计量分析 [J]. 湖北中医药大学学报, 2015, 17 (3): 113 - 116.
- 2 Murdoch T B, Detsky A S. The Inevitable Application of Big Data to Health Care [J]. JAMA, 2013, 309 (13): 1351 - 1352.
- 3 李芳, 陈颖, 侯丽, 等. RxNorm 多词表语义互操作机制研究 [J]. 数字图书馆论坛, 2012, (12): 65 - 71.
- 4 Deng ZH, Tang SW, Zhang M, et al. Overview of Ontology [J]. ACTA Scientiarum Naturalum Universitatis Pekinesis, 2002, (38): 730 - 738.
- 5 马利, 崔志伟, 毛树松. 我国医学知识库应用现状研究 [J]. 医学信息学杂志, 2013, 34 (11): 55 - 59.
- 6 徐彬锋, 温志浩, 罗小刚, 等. 基于本体的医学知识库构建及应用 [J]. 北京生物医学工程, 2011, 30 (6):

618 - 623.

- 7 周肖彬, 曹存根. 基于本体的医学知识获取 [J]. 计算机科学, 2003, 30 (10): 35 - 39, 54.
- 8 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.
- 9 巩泳歌, 温有奎. 基于本体的高血压疾病诊断知识库 [J]. 情报杂志, 2010, 19 (6): 169 - 172.
- 10 李新霞. 基于本体的中医学脾胃病知识库的构建 [D]. 南京: 南京理工大学, 2008.
- 11 冯贞贞. 基于本体的临床路径知识库的研究与构建 [D]. 苏州: 苏州大学, 2012.
- 12 罗雪琼, 周毅, 胡珊, 等. 泌尿外科本体知识库构建研究 [J]. 中国数字医学, 2015, 10 (5): 27 - 30.
- 13 苟玲, 张艺帆, 陆遥, 等. 基于本体的抗生素临床管理知识库的建立 [J]. 中国数字医学, 2015, 10 (10): 88 - 91.
- 14 陈布伟, 李冠宇, 张俊, 等. 基于语义网规则语言的推理机制框架设计 [J]. 计算机工程与设计, 2010, 31 (4): 847 - 849, 853.
- 15 徐德智, 汪智勇, 王斌. 当前主要本体推理工具的比较分析与研究 [J]. 现代图书情报技术, 2006, (12): 12 - 15, 77.
- 16 Horst H T. Completeness, Decidability and Complexity of Entailment for RDF Schema and a Semantic Extension Involving the OWL Vocabulary [J]. WEB Semantics Science Services & Agents on the WO, 2005, 3 (2/3): 79 - 115.
- 17 Friedman E. Jess, the Rule Engine for the Java Platform [J]. Hill, 2006, 48 (2): 340 - 354.
- 18 Eeiksson H. Using Jesstab to Integrate Protégé and Jess [J]. Intelligent Systems IEEE, 2003, 18 (2): 43 - 50.
- 19 范轶. 基于本体推理的心电图辅助诊断系统研究 [D]. 长春: 吉林大学, 2010.
- 20 肖敏. 基于本体推理的健康监测平台的研究与实现 [D]. 长春: 吉林大学, 2013.
- 21 许德山, 乔晓东, 朱礼军, 等. 本体推理在知识检索中的应用 [J]. 现代图书情报技术, 2009, (1): 58 - 63.
- 22 李跃新, 胡婕, 秦丽, 等. 知识工程基础与应用案例 [M]. 北京: 科学出版社, 2006: 100 - 101.
- 23 刘林, 周雪忠, 周霞继, 等. 国际临床表型的本体研究概述及问题探讨 [J]. 世界科学技术 - 中医药现代化, 2015, (8): 1634 - 1638.
- 24 宣腾. 区域医疗本体知识库构建及其语义应用 [D]. 成都: 电子科技大学, 2013.