

基于领域本体的中医知识获取方法

蒋宏潮, 王大亮, 张德政

(北京科技大学信息工程学院, 北京 100083)

摘要: 中医知识获取是计算机自动实现知识管理、知识重组的前提。中医领域存在大量非结构化知识, 难以被深入获取。该文提出一种基于领域本体的知识获取方法, 对选取的单一医案中蕴含的知识进行形式化表示, 构造中医领域的语义网络, 对其进行定量分析, 抽取中医领域的概念语义场。实验结果表明, 该方法可在一定程度实现中医领域的知识获取。

关键词: 本体; 中医; 语义场; 知识获取

Approach of Chinese Medicine Knowledge Acquisition Based on Domain Ontology

JIANG Hong-chao, WANG Da-liang, ZHANG De-zheng

(School of Information Engineering, University of Science and Technology, Beijing 100083)

【Abstract】 Acquiring knowledge of Chinese medicine is the premise of knowledge management and knowledge integration by computer. There's lots of unstructured knowledge in Chinese medicine domain, which is difficult to get. An approach of knowledge acquisition based on domain ontology is presented to obtain the formal representation of knowledge in a single medical selected, structure the semantic network, analyze it quantitatively, and acquire the concept semantic field on Chinese medicine domain. The experimental results show that the approach can get Chinese medicine domain knowledge.

【Key words】 ontology; Chinese medicine; semantic field; knowledge acquiring

中医是通过传承、实践和创新而形成的独特知识体系。在中医辨证施治过程和医案中, 蕴含了丰富的、独具特色的学术思想、诊断经验和治疗方法。从中医知识的形式化表示中获取领域知识, 是实现计算机自动完成知识重组、知识管理及重现中医专家诊疗规律的关键。

1 领域本体及其构建技术

本体论是客观存在的一个系统的解释或说明, 关心的是客观现实的抽象本质。在人工智能界, Neches等人最早对本体论定义如下: 给出构成相关领域词汇的基本术语和关系, 以及利用这些术语和关系构成的规定这些词汇外延的规则的定义^[1]。

根据本体论观点, 中医是由概念及概念之间的关系构成的知识体系。因此, 可以为中医构造一个形式化的领域本体系统。

1.1 本体建模原语

Perez 用分类法组织本体归纳出如下 5 个基本建模元语:

(1)类或概念。指任何事务, 如工作描述、功能、行为、策略和推理过程。从语义上讲, 它表示的是对象的集合, 其定义一般采用框架结构, 包括概念的名称, 与其他概念之间的关系的关系的集合, 以及用自然语言对概念的描述。

(2)关系。在领域中概念之间的交互作用, 形式化的定义为 n 维笛卡儿积的子集, $R: C_1 \times C_2 \times \dots \times C_n$, 如子类关系 (subclass-of), 在语义上关系对应于对象元组的集合。

(3)函数。函数关系的前 $n-1$ 个元素可以唯一决定第 n 个元素。形式化的定义为 $F: C_1 \times C_2 \times \dots \times C_{n-1} \rightarrow C_n$, 如 mother-of 是一个函数, mother-of(x, y)表示 y 是 x 的母亲。

(4)公理。代表永真断言, 如概念乙属于概念甲的范围。

(5)实例。代表元素, 从语义上讲实例表示的就是对象。

按照 Tom Gruber 给出的评价本体设计的基本原则, 还须满足客观性、完全性、一致性、最大单调可扩展性、最小承诺和最小编码承诺。

1.2 本体建模语言

图 1 表现了一个中医领域的有向概念。

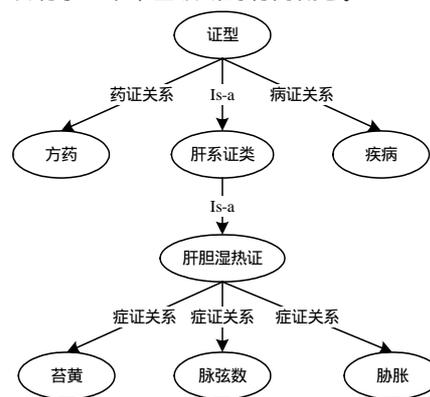


图 1 中医领域的有向概念

本体有两个基本要素: 概念和概念之间的关系。可以用一个二元组 $Onto(O, R)$ 来表示本体, 其中, O 表示领域的概念集, 它包括领域内的类、类的属性及类的实例, 类是概念

基金项目: 国家“863”计划基金资助项目(2007AA01Z170)

作者简介: 蒋宏潮(1975 -), 男, 博士研究生, 主研方向: 自然语言处理; 王大亮, 博士研究生; 张德政, 教授、博士

收稿日期: 2007-08-10 **E-mail:** lizju196@163.com

的核心，类属性及实例都是类的扩展； R 表示领域内概念的关系集合。

可以使用一个有向的概念图来表示本体，其中的顶点表示概念，有向边表示概念间的关系^[2]。

1.3 中医领域本体的构建方法

由于对各自学科领域和具体工程的不同，因此构建知识本体的方法各不相同。目前尚没有一套标准的知识本体构建方法，但是构造本体时必须遵守一些原则。本文构建的中医领域本体符合 Gruber 于 1995 年提出的以下 5 条规则：

- (1)明确性和客观性。使用自然语言对术语给出明确、客观的语义定义。
- (2)完整性。给出的定义是完整的，能表达特定术语含义。
- (3)一致性。知识推理产生的结论与术语本身的含义不会产生矛盾。
- (4)最大单项可扩展性。向本体中添加通用或专用的术语时，通常无须修改已有内容。
- (5)最少约束。尽可能少对建模对象列出约束限定条件。

中医领域存在大量的非结构化知识，本文建立中医领域本体的过程，使用图 2 进行描述。并使用了 protégé3.2 工具完成构建中医本体。

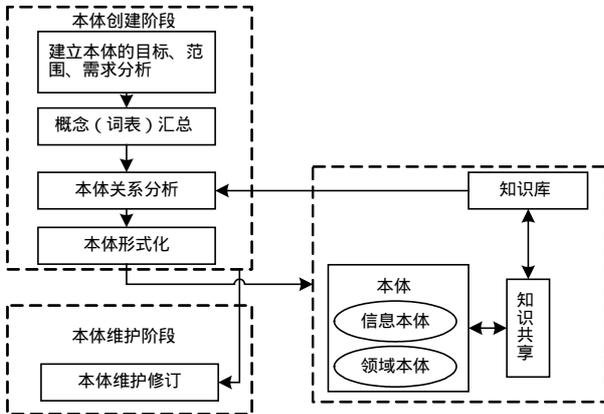


图 2 中医领域本体的建立与应用

2 中医知识获取

知识获取就是把用于求解专门领域问题的知识从拥有这些知识的知识源中抽取出来，并转换为一种特定的计算机表示。

2.1 知识获取的难点

获取中医领域专家的启发性知识难度很大，原因如下：

- (1)知识表示不一致。专家通常陈述知识的方法与专家系统采用的知识表示方法不一致。
- (2)专家的启发性知识不够精确。专家的启发性知识往往隐含着近似、不确定、不充分、不完全，甚至产生矛盾。
- (3)有些启发性知识表示的不可能性。领域专家凭借多年总结和积累的实践经验，采用独特的方法和有效的手段去解决困难问题，但难以把这些经验和策略方法显式地表达出来。
- (4)缺乏开发专家系统的现代技术。现行系统采用的表示方法限制了它的表达能力。
- (5)知识测试与调试的困难性。知识的正确性需要经过反复测试与调试，为了孤立出形成问题解答的错误，可能需要跟踪包含着数百个事实的几十种推理。为了使观察到的错误与它的真实原因联系起来，必须弄清知识与推理机控制策略之间的相互作用。而且，除非知识各部分之间相互依赖关系是非常明了的，否则，在修正一个观察到的错误时，在知识库中的修改都可能引起新的错误，这些错误有可能降低系统

的性能。

2.2 语义网

语义网将人的记忆模型画为一个联系网，将其中的节点用来表示实体、概念，而连接则表示概念间的联系^[3-4]。根据中医理论体系知识结构的特点，为了完成中医专家系统的构造，本文采用语义网络来表示中医诊疗知识，并用于构建中医知识库。

(1)常用语义网

在语义网表示中，节点可以划分为实例节点和类节点两种类型，节点间的弧用于表示节点之间的联系。在表达事实性知识和事实之间联系的知识时，语义网采用统一的形式。

- 1)ISA 类型，用于表示实例——类属关系。
- 2)AKO 类型，全称为 A-KIND-OF，用于表示具体类——抽象类关系。ISA 将某个具体实例与一个一般类联系起来，AKO 则将类与类联系起来。
- 3)IS 类型，IS 弧用于表示某一节点是另一节点的属性。
- 4)PART-OF 类型，这种类型用于表示部分与整体的关系。它基于概念的分解，通常将高层次的概念分解为若干低层次概念的集合。

(2)事件知识的表示

设有 n 元谓词或关系 $R(X_1, X_2, \dots, X_n)$ ，其中 X_i 的取值为 $a_i (1 \leq i \leq n)$ 。可以将 R 转化为等价的一组 n 个二元关系：

- $$\begin{aligned} X_1 (R, a_1) \\ X_2 (R, a_2) \\ \dots \\ X_n (R, a_n) \end{aligned}$$

将 R 和具体值 a_i 作为语义节点，将 X_i 作为节点 R 和节点 a_i 之间的弧。这样就可以用语义网表示事件知识。比如“手太阴肺经循行由胸走手”，谓词形式为经络循行(手太阴肺经，胸，手)，即可以表示为如图 3 所示的语义网。

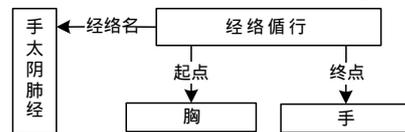


图 3 使用语义网表示中医事件

(3)联结词在语义网中的表示方法

语义网具有与谓词逻辑等等的表达能力，它不仅可以表示基本命题，还可以表示与、或、非以及蕴涵关系。

在表示各种基本命题的组合方式时，只需要引入相应的组合节点就可以了，比如“与”、“或”节点。

对于否定类型的关系，可以直接采用带有否定标志的弧来表示，比如，为表示事实“当归不是泻下药”，可以在节点“当归”和节点“泻下药”之间使用~AKO 弧来表示。

对于蕴涵关系，本文引入一个新的蕴涵节点表示规则中前提条件和结论之间的因果关系。从蕴涵节点引出两条弧。一条记为 ANTE，指向前提条件，另一条记为 CONSE，指向结论。

例如，对于规则 $\sim P \rightarrow Q \rightarrow R \rightarrow S$ ，可以用图 4 所示的语义网来表示。



图 4 规则 $\sim P \rightarrow Q \rightarrow R \rightarrow S$ 的语义网表示

