

基于领域本体的方剂知识获取与研究

马 斌, 柴 智

(山西中医学院 医药管理学院, 山西 太原 030024)

摘 要: 中医方剂知识获取是计算机实现自动知识管理的前提。计算机本体理论技术是构建专业领域模型的主要方法, 能够有效挖掘中医药领域存在的大量隐性知识。文中以中医方剂为切入点, 依照中医学科体系特点, 设计语义与学科相结合的分类体系与概念关系, 明确描述方剂基本概念。提出一种基于领域本体的知识构建方法, 对选取的方剂知识进行形式化表示, 并运用 Protégé 完成了清热解毒方本体的实现。实验结果表明, 该方法可以在一定程度实现中医领域的知识获取。

关键词: 知识获取; 领域本体; 方剂研究; Protégé

中图分类号: TP182

文献标识码: A

文章编号: 1673-629X(2013)06-0227-03

doi: 10.3969/j.issn.1673-629X.2013.06.058

Chinese Medicine Prescription Knowledge Acquisition and Research Based on Domain Ontology

MA Bin, CHAI Zhi

(College of Medical Management, Shanxi College of Traditional Chinese Medicine, Taiyuan 030024, China)

Abstract: Acquiring knowledge of Chinese medicine prescription is the premise of knowledge management by computer. Ontology theory is a new method in constructing professional domain model, it can be effective mining field existing large amount of tacit knowledge. Focus on Chinese medicine prescription, according to traditional Chinese medicine discipline system characteristics, design classification system and concept relations of combination of semantic and subject, clearly define the basic concept of prescriptions. An approach of knowledge construction based on domain ontology is presented to obtain the formal representation of prescription knowledge selected, and finish the ontology of clearing heat and removing toxicity by Protégé. The experimental results show that the approach can get Chinese medicine domain knowledge to some extent.

Key words: knowledge acquisition; domain ontology; prescription research; Protégé

0 引 言

本体(Ontology)一种新的知识模型表示方法和共享载体,已成为自然语言理解、知识工程、信息系统集成、信息抽取与检索等多个领域的研究热点。本体是对某一领域内共同认可的概念及概念之间关系的显式说明和形式化表示,它能够较好地建立用户和计算机之间基于语义的信息交流,并为操作者及广泛异构的应用系统之间提供统一的领域知识理解。本体能有效将大量复杂的数据、知识等资源进行管理,强调知识的重用与共享,是目前构建领域知识库的重要方法。

我国传统中医药经历了数千年的发展,形成了完备的知识体系结构,具有重要的学术价值和实用价值。

但随着信息化的深入发展和知识发现等概念的提出,其传统的知识组织方式已经不能适应现代社会信息的需求,在一定程度上阻碍了中医药知识的有效利用。将计算机领域先进的本体理论与技术引入到中医药知识组织研究中,有利于克服中医药概念的不规范性与语义屏障,实现中医药信息的知识化重组,为中药领域隐性知识的挖掘和利用提供有力保障。

1 本体基本概述

1.1 本体定义

本体概念最早出现于哲学领域^[1],是关于世界某个方面特定的分类体系,这个系统不依赖于任何一种

收稿日期:2012-09-11

修回日期:2012-12-25

网络出版时间:2013-03-05

基金项目:山西省青年科技研究基金(2012021031-6)

作者简介:马 斌(1979-),男,山西太原人,讲师,硕士,CCF会员,研究方向为数据挖掘与人工智能。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20130305.0816.019.html>

特定的描述语言,反映的是客观现实的抽象本质。1993年,Gruber首先给出了一个最为流行的定义,即“本体是概念化的明确的规范说明”^[2]。后来随着在知识工程、人工智能、信息检索与利用等领域中的广泛应用,其定义也被融入了许多新的内涵。Borst认为:“本体是共享概念模型的形式化规范说明”^[3]。1998年德国学者 Studer 对上述两种概念进行分析,将本体定义为“共享概念模型的明确的形式化规范说明”^[4],从知识管理角度分析,本体提供了知识领域的共同理解,并从不同层次给出这些词汇间相互关系的明确定义,减少了人们对问题域中概念处理和逻辑关系之间存在的误解。

1.2 本体建模元语

本体常用建模元语有:

1)类(Classes):表示对象元素的集合,采用框架结构来定义其内涵,包括概念的名称、相互关系和自然语言对概念的描述。

2)函数(Functions):指一类特殊的关系。该关系的前 $n-1$ 个元素可以唯一决定第 n 个元素。

3)关系(Relations):指领域中概念之间的相互作用,用 N 维笛卡儿积的子集表示为 $R: C_1 \times C_2 \times \dots \times C_n$ 。

4)公理(Axioms):公理代表永真断言,如概念 A 属于概念 B 的范围。

1.3 本体描述语言和开发工具

本体描述语言来源于人工智能领域对知识管理的分析研究,它协助用户为领域模型编写形式化的、清晰的概念描述。本体描述语言不但具有规范的语法、语义和充分的表达方式,更有完备的推理支持能力^[5]。目前使用最普遍的本体描述语言有 Ontolingua、RDF(S)、OWL、Loom 等^[6]。OWL 是 W3C 建立的国际通用的标准本体描述语言,是网络环境下发布和共享本体的语义标记语言^[7]。OWL 建立在 RDF(S) 等已有标准基础上,通过添加大量的基于描述逻辑的语义原语来表示各种本体^[8]。文中采用 OWL 语言对中医方剂本体内容进行规范组织。

本体开发工具有编辑、图示、自动转换置标语言、附加软件等功能。目前其种类主要有:Protégé、OntoEdit、OilEd、Open Cyc 等。本研究使用 Stanford 大学医学部的 Protégé 作为本体开发工具,它采用 Java 语言编写,提供一种可扩展的结构框架,模块划分清晰,使用简单方便,可以很好地用于中医方剂领域本体。

2 方剂领域本体的构建

2.1 方剂本体构建方法

在实际构建方剂本体时,设计人员应对所涉及的

专业领域有一定的知识储备,熟悉中医方剂专业基础知识内容和本体开发技术,能设计出完整的概念关系和逻辑结构,并建立基本语义模型。构建相应的本体实施方案是建立大型领域本体的第一步,对于知识库的组织管理起着至关重要的作用。通过上述分析,本研究在斯坦福大学“101 方法”^[9]的基础之上,提出一种基于本体思想的由方剂领域专家参与的方剂知识本体构建方法。具体流程见图 1。

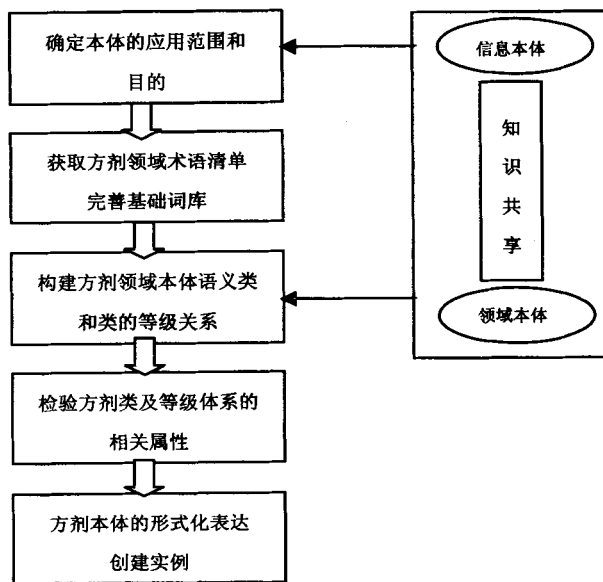


图 1 方剂领域本体构建流程

2.2 方剂语义网知识表示

语义网用节点来表示实体和概念,节点分为类节点和实例节点,节点之间的联系用弧来描述^[10]。由于弧具有方向性,在弧上可以标注信息来表示节点之间的联系。构建语义网的目标是获取和处理相关领域知识,并在此基础上进行深入分析和推理。

2.2.1 语义网基本类型

1)PART-OF 类型,表达概念整体与部分之间的联系,它是基于概念的分解,通常将高层次的概念分解为若干低层次概念的集合。

2)KIND-OF 类型,用于表达概念之间的继承关系。

3)INSTANCE-OF,表达概念实例与概念之间的关系。

4)ATTRIBUTE-OF,表达某个概念是另一个概念的属性。如“组成”是方剂的一个属性。

2.2.2 方剂知识表示

方剂概念之间的关系通过形成 RDF 文件来存储相关的信息。方剂本体的实现可以运用 RDF 的资源、属性及属性值三者相对应的形式存储所描述的的资源信息,这样 RDF 就构成一个完整的知识描述文件。

例如以“黄连解毒汤”为例,其部分概念表达的 RDF 文件如下:

```

<rdf:RDF xml ns:rdf=" & rdf "
Xml ns: kb=" & kb"
Xml ns: rdfs=" & rdfs" >
<kb:麻黄汤 rdf: about=" & kb;KB_508159_Instance_4"
kb:方名=" 黄连解毒汤"
kb:来源="《外台秘要》引崔氏方"
kb:用法="上四味切,以水六升,煮取二升,分二服"
rdfs:label=" KB_508159_Instance_4" >
<kb:方剂功用 rdf:about=" & kb;方剂_Instance_5"
kb:功用主名=" 发汗解表"
rdfs:label=" 方剂_Instance_5" >
<kb:功用相关概念词>泻火解毒</kb:功用相关概念词>
</kb:方剂功用>
</kb:方剂组成>
</rdf:RDF>

```

3 实验结果

3.1 创建类及命名类

类的层次设计是指不同专指度的上位概念与下位概念之间的关系,上位概念常常是下位概念的抽象、概括或整体表示;下位概念往往是对上位概念的补充和细化。在方剂的上下位概念层次里标明了方剂的分类关系,形成了方剂的分类表,为查询方剂相关信息提供了快捷的途径。按照方剂的功能分类设计的概念的层次关系如图2所示:

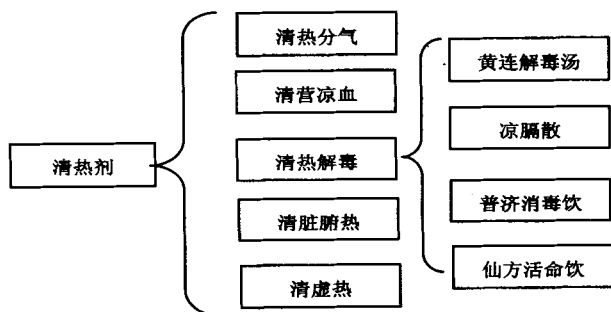


图2 清热解毒方类的层次

Protégé 中的 OWL/RDF Files 文件具有多插件支持和在本地编辑工具中进行修改等特点。鉴于这些优点,本研究选择 OWL/RDF Files 作为基本的文件类型来建立类目,在 OWL Class 标签下, SUBCLASS EXPLORER 面板中“Thing”为系统默认存在的顶级类,所有的客观实体都是它的子类,在顶级类“Thing”的菜单中选择“Create subclass”来建立新的下位类。在创建“中医方剂”新类后,在 CLASS EDITOR 中可以编辑此类。

3.2 建立对应属性

通过属性参数,从各个角度尽可能全面地描述方剂,可以更精确、全面地显示概念间的关系。在属性设置时,我们主要定义方剂类的方名、组成、用法、功用、主治、方解等六个基本属性。在 Protégé 中 PROPERTY

BROWSER 环境下通过 Annotation、Object、Data type 属性的设置来对应方剂不同的关系,实现类目属性的操作。

3.3 编辑方剂实例

在 Protégé 中,通过 Individuals 标签的来添加方剂实例。内容包括实例名称及附加在该类上的属性。当大量的实例被赋予相应的属性及属性值时,就形成了相关领域的知识库。实例可以自动继承类的属性设置,方剂本体主要对下位类概念进行了实例导入。

4 清热解毒方本体的可视化表达

本体的可视化表现有助于丰富本体的概念和关系的展示,提高用户的认知水平,促进知识共享。此外,用户在相关领域的本体构建中可以方便的了解到本体的概念体系,有利于本体的复用和修正。目前,本体的编辑工具 Protégé 中包含了很多可视化插件,用户可以利用其中的相关插件进行图形化的浏览,如 TGViz-Tab、OntoViz、Jambalaya 等,这些操作工具在本体的编辑过程中是很实用的。

5 结束语

文中探讨了中医药方剂领域知识表示的一种规范形式,研究了方剂领域本体结构,提供了计算机处理方剂知识可理解、可共享、可复用的数据模式,为进一步构建中医领域的本体系统、挖掘整理中医临证经验及建立基于信息检索平台的中医知识库打下了基础。但是中医方剂知识库的建立是不断推敲和完善的过程,我们需要在应用中不断进行修改与扩充。例如对中医药方剂知识的动态描述和方剂配伍规律的进一步认识,都需请相关专家介入以提高实际操作的质量,以期实现中医领域本体的自动化、智能化应用。

参考文献:

- [1] Towards R G. Principles for the Design of Ontologies Used for Knowledge Sharing[J]. International Journal of Human-computer Studies, 1995, 43: 907-928.
- [2] Gruber T R. A translation approach to portable ontology specifications[R]. [s. l.]: Knowledge System Laboratory, 1993.
- [3] 刘燕玲, 华庆一, 郭晓娟. 基于领域本体面向问题的需求分析与领域建模[J]. 计算机技术与发展, 2007, 17(8): 99-102.
- [4] Rudi S, Benjamins R, Fensel D. Knowledge Engineering: Principles and Methods[J]. Data & Knowledge Engineering, 1998, 25(1/2): 161-197.
- [5] 曹宇峰, 曹存根. 基于本体的中医舌诊知识的获取[J]. 计算机应用研究, 2006(3): 31-34.

(下转第 232 页)

基于领域本体的方剂知识获取与研究

作者: [马斌, 柴智, MA Bin, CHAI Zhi](#)
作者单位: [山西中医学院医药管理学院, 山西太原, 030024](#)
刊名: [计算机技术与发展](#) 
英文刊名: [Computer Technology and Development](#)
年, 卷(期): 2013, 23(6)

本文链接: http://d.g.wanfangdata.com.cn/Periodical_wjz201306058.aspx