

· 学术探讨 ·

中医哮喘领域本体的构建

朱玲¹, 李敬华¹, 于琦¹, 于彤¹, 田野¹, 孙晓峰¹, 徐丽丽¹, 李泽庚^{2*}

(1. 中国中医科学院 中医药信息研究所, 北京 100700;

2. 安徽中医药大学 第一附属医院, 合肥 230038)

[摘要] **目的:**利用本体的方式实现中医哮喘领域的知识的共建共享及推理。**方法:**通过对中医文献中“哮喘”相关知识的提取,基于斯坦福大学的七步法构建了“哮喘”的疾病本体,实现了哮喘相关知识的领域建模,建立了部分概念与概念间的关系约束及公理。**结果:**中医哮喘领域本体构建完成能够实现知识的共享以及简单的推理。共有类120个,实例425个,对象属性18个,其中反向属性6个,数值属性2个,等价类公理1个,子类公理138个,基本实现了对哮喘领域知识的语义化表达。**结论:**本体的构建在知识的共建、共享及未来知识库的语义检索、数据挖掘中具有一定的参考与应用价值。

[关键词] 哮喘; 本体; 知识库; 数值属性; 语义网络; 等价类公理

[中图分类号] R256.12; B016; R24 **[文献标识码]** A **[文章编号]** 1005-9903(2017)15-0222-05

[doi] 10.13422/j.cnki.syfjx.2017150222

[网络出版地址] <http://kns.cnki.net/kcms/detail/11.3495.R.20170511.0900.008.html>

[网络出版时间] 2017-05-11 9:00

Construction of Ontology of Asthma in Traditional Chinese Medicine

ZHU Ling¹, LI Jing-hua¹, YU Qi¹, YU Tong¹, TIAN Ye¹, SUN Xiao-feng¹, XU Li-li¹, LI Ze-geng^{2*}

(1. *Institute of Information on Traditional Chinese Medicine, China Academy of Chinese Medical Sciences, Beijing 100700, China;*

2. *The First Affiliated Hospital of Anhui University of Chinese Medicine, Hefei 230038, China*)

[Abstract] **Objective:** To realize the sharing and reasoning of knowledge in the field of Chinese medicine asthma by ontology. **Method:** Based on seven-step method from Stanford university, the disease ontology of asthma which included axioms was constructed by realizing the knowledge of asthma-related knowledge, and the relationship between some concepts. **Result:** The ontology construction of Chinese medicine asthma field was completed, which could realize the sharing of knowledge and simple reasoning, containing 102 classes, 425 individuals, 18 object properties, 6 inverse properties, 2 data properties, 1 equivalent class axiom and 138 subclass of axioms counts. **Conclusion:** The construction of ontology has a certain value in the co-construction of knowledge and the semantic retrieval and data mining of the future knowledge base.

[Key words] asthma; ontology; knowledge base; numerical attribute; semantic network; equivalent class axiom

支气管哮喘简称哮喘,主要是由T淋巴细胞、肥大细胞和嗜酸性粒细胞浸润而导致的一种常见的慢性呼吸道变态反应炎症。世界卫生组织估计,

全球约有3亿哮喘患者,其中每年约有1500万人因为该病而丧失劳动能力,目前中国已有2000万的哮喘患者^[1],给社会带来了严重的经济负担,给国民

[收稿日期] 20170208(004)

[基金项目] 国家中医药管理局中医防治重大疑难疾病能力建设项目(2014)

[第一作者] 朱玲,副研究员,博士,从事中医药信息研究,Tel:13426395263,E-mail:jjzhuling@163.com

[通讯作者] *李泽庚,教授,从事中医药防治呼吸系统疾病的临床研究,Tel:13805516609,E-mail:13805516609@126.com

造成了巨大的经济损失,已成为严重的公共卫生问题之一。

研究哮喘患病率及相关因素、构建相应的知识库及疾病诊断专家系统将有助于为制定合理的防治策略提供依据,从而达到降低疾病负担、提升国民健康的期望。但由于应用目的及用户需求不同,这些系统采用了不同的知识表示、建模方法及开发工具来解决不同的领域问题,常常会出现同词不同义等问题,从而造成对概念理解的分歧,进而导致系统无法共享及交互操作。

本体^[2]作为一种知识建模工具,自被提出以来就引起了国内外众多科研人员的广泛关注。由于本体能够很好地描述概念以及概念与概念之间的关系,具有良好的概念层次结构和对逻辑推理的支持,因而本体的提出可以为概念理解的歧义提供解决方案,进而为知识的共享提供可能。杨春媛^[3]详尽而系统地介绍了生物医学领域中的人类肝脏疾病本体,在肝脏疾病相关知识和分子相互作用信息的挖掘、整合、交流等方面具有着基础性的意义。此外,还可以基于疾病本体(disease ontology, DO)及其注释信息进行疾病相似性计算,对于复杂疾病发病机制的理解、诊断、预测和药物研发具有重要意义。

随着本体研究的深入,其实践工具逐渐成熟,主要包括 Protégé, OntoEdit, OilEd, RDF Gateway 等,其中 Protégé 是目前比较流行的开源软件。允许用户在概念层次上进行设计,不需要了解具体的本体描述语言,就可以非常方便地构造本体模型,提供文本,HTML, JDBC, RDF Schema, XML Schema 等多种输出格式,同时提供完全的应用程序编程接口(API),支持插件扩展更多特殊功能。该软件由美国斯坦福大学采用 Java 开发,目前版本为 Protégé 5.1,但有鉴于 Protégé 5.1 在中文支持方面的缺陷,故选择中文显示更为流畅的 Protégé 4.1 作为本研究的工具。本研究拟以哮喘知识体系为研究对象,在系统考辨哮喘疾病相关概念术语,明确其内涵外延、演变、相互关系的基础上,以本体论方法为指导,运用医学信息学技术建立其知识源数据,以 Protégé 4.1 为工具,以网络本体语言(OWL)为描述语言,构建哮喘领域本体,展现哮喘相关概念之间的逻辑关系,揭示哮喘领域概念之间的隐性关系,从而有利于相关文献的全面检索、数据库建设、文本挖掘。

1 领域本体的建立

1.1 获取领域中的重要知识 本体的创建流程见图 1。从医学主题词表(medical subject headings,

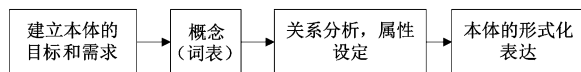


图 1 本体的创建流程

Fig. 1 Creation process of ontology

MeSH), 中文医学主题词表, 中医临床术语系统^[4], 内科教材, 辞典, 相关标准及其他相关专科疾病书籍^[5]等文献中筛选出重要专业术语, 收集所有有用的和潜在有用的相关领域专家认可的概念、语义、属性、实例等, 再经过整理、提炼, 从而建立术语汇总表, 并将术语的解释一并收入, 力求概念集的完整性, 暂不考虑这些术语的概念重叠、关联性及概念属性。

1.2 确定基本概念及其关系阶段 上一步骤中已经产生了大量的概念, 见图 2, 但却缺乏组织, 这需要按照一定的规则将其分组, 形成不同的类别。同一类别的概念, 其相关性相对较强。

另外, 需对每一个概念的重要性进行评估, 选出关键性术语, 尽可能准确而精简地表达出领域的知识。以医学系统命名法-临床术语^[6](systematized nomenclature of medicine-clinical terms, SNOMED-CT)为参照, 中医临床术语为架构, 疾病哮喘为切入点, 形成一个领域知识的框架体系, 参考中医药学语言系统^[7], 厘清领域知识内常用的语义网络, 见图 3, 建立语义关联三元组, 为领域本体的框架结构提供支撑。

1.3 本体指标 哮喘疾病本体共有类(class)120 个, 实例(individual)425 个, 对象属性(object property)18 个, 其中反向属性(inverse property)6 个, 数值属性(data property)2 个, 等价类公理(equivalent class axiom)1 个, 子类公理(subclass of axioms count)138 个, 基本实现了对哮喘领域知识的语义化表达。

1.4 属性的设置

1.4.1 对象属性 属性的设置中最为重要、使用最为频繁的是对象属性的设置, 对象属性主要用来描述 2 个实体之间的关系, 就该领域而言, 哮喘本体包括时间上相关的“与……同时发生”, 空间上相关的“位于”, 功能上相关的“使用”、“作用于”、“导致”、“擅长”、“现象表达”、“是……现象表达”、“治疗”、“被……治疗”、“组成”、“由……组成”、“使用”、“被……使用”、“预防”、“被……预防”、“禁止使用”等。

对象属性的特征建立有助于推理机提供更多可供理解的推论。如果一个属性可以有传递性, 举例

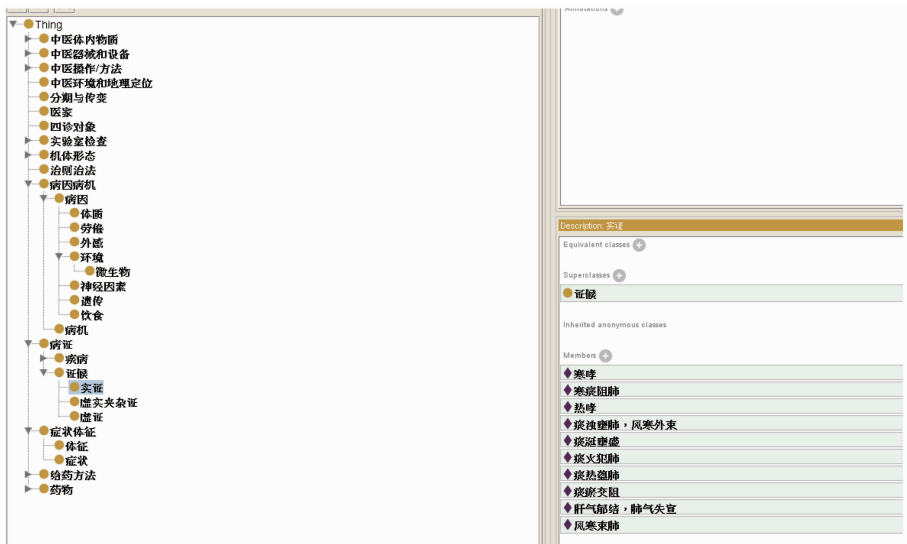


图 2 哮喘证候实例
Fig. 2 Examples of asthma syndrome

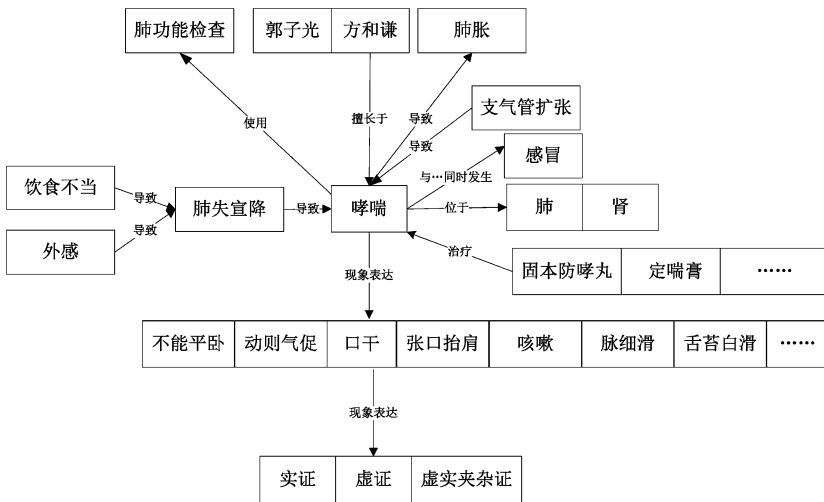


图 3 哮喘的语义网络
Fig. 3 Semantic network of asthma

来说,如果一个家庭成员 1 是 2 的兄弟,2 是 3 的兄弟,然后就可以推论出 3 是 1 的兄弟。兄弟之间的关系就是可以传递的。通过对对象属性 domain 和 range 的限定,可以规范未来的数据质量,比如“由……组成”的 domain 是中成药或方剂,range 是中药或西药,“治疗”的 domain 是中医操作/方法或药物或方剂或中成药,range 是病证或症状体征或证候。“现象表达”的 domain 是疾病或病证,range 是症状体征。如果属性相关的 domain 和 range 设置错误,则本体推理机就会提示存在问题,为保证本体的内在一致性奠定基础。

1.4.2 数值属性 数值属性用于描述个体与数值之间的关系,比如某人的年龄 18 岁,身高 158 cm,体重 55 kg 皆为数值属性。数值属性的设置在现有

本体表达中较少,比如美国密歇根大学开发的不良反应本体(OAE)^[8]也仅有 1 种,基于疾病哮喘所涉及的具体属性,本文设立 2 种数值属性——“有最大呼气流速”、“有肺活量”,通过对肺功能的检查可以判断病情的严重程度,为疗效的判断和预后提供参考。比如哮喘发作程度较轻的患者,最大呼气流速 $> 200 \text{ L} \cdot \text{min}^{-1}$ 。

这样的知识表示,面对具体病人的肺功能测试结果,计算机就可以很容易判断其属于哮喘轻证还是重证。如果肺功能显示肺活量 $> 2 \text{ L}$,最大呼气流速 $> 200 \text{ L} \cdot \text{min}^{-1}$,则提示哮喘轻证,如果 1 s 内呼气肺活量 $< 1 \text{ L}$,最大呼气流速 $< 80 \text{ L} \cdot \text{min}^{-1}$,则提示哮喘重证,肺功能的检测可以为疗效的判断和预后提供参考。关于数值属性,未来可以进一步扩充,比



图 8 本体的推理示例

Fig. 8 Reasoning example of ontology

4 讨论

本研究在正确理解、解释概念术语内涵及外延的基础上,构建了哮喘疾病本体,对哮喘领域知识体系进行重新组织和表示,不但可以更科学、更有效地发现、管理、利用相关知识,提高信息的传递速率与理解深度,还有利于哮喘文献的全面检索、数据库建设、文本挖掘以及相关概念的正确阐释,以便将来构建整个哮喘疾病得语义网络,将本体应用于临床医学领域,将有利于中医临床术语系统的优化与完善,这也在一定程度上丰富了本体语言,对未来构建更加复杂和完善的哮喘临床管理知识库具有一定的参考作用。

由于时间限制,哮喘本体库在完整性、合理性及运行性方面依然存在一些不足,未来应扩增哮喘的概念术语,增加哮喘领域的公理表示,建立以本体为基础的结构清晰、操作简洁的知识库,通过本体推理实现诊断与治疗用药的推荐。未来若可以将本体库

与电子病历系统进行动态交互,则在临床医学领域的应用将更为广泛和深入,为新一代临床信息系统的发展奠定基础。

[参考文献]

- [1] 冯晓凯. 我国支气管哮喘患病情况及相关危险因素流行病学调查[D]. 北京:北京协和医学院,2014.
- [2] 杨月华,杜军平,平原. 基于本体的智能信息检索系统[J]. 软件学报,2015,26(7):1675-1687.
- [3] 杨春媛. 人类肝脏疾病本体的构建及其应用[D]. 北京:北京协和医学院,2012.
- [4] 中国中医科学院中医药信息研究所. 中医临床术语网[EB/OL]. http://192.168.200.195:8080/tpweb/CommonAction_main,2017-04-11.
- [5] 冯维斌,刘伟胜. 呼吸科专病中医临床诊治[M]. 2版. 北京:人民卫生出版社,2005:2-434.
- [6] International health terminology standards development organization. SNOMED [EB/OL]. <http://browser.ihtsdotools.org/?perspective=full&conceptId1=108828006&edition=en-edition&release=v20160131&server=http://browser.ihtsdotools.org/api/snomed&langRefset=900000000000509007>,2015-07-19.
- [7] 于彤,崔蒙,李海燕,等. 中医药学语言系统的语义网络框架:一个面向中医药领域的规范化顶层本体[J]. 中国数字医学,2014,9(1):44-47.
- [8] HE Y, Sarntivijai S, YU L, et al. OAE: the ontology of adverse events [J]. J Biomed Semantics, 2014, doi: 10.1186/2041-1480-5-29.

[责任编辑 刘德文]