

语义 Web 在中医药领域的应用研究综述

于 彤 崔 蒙 李敬华

(中国中医科学院中医药信息研究所, 北京, 100700)

摘要 语义 Web 是一项新兴的 Web 技术,用以构建一个全球性的巨型数据网络,从而支持更加智能的 Web 应用。中医团体在语义 Web 的推广方面,进行了一系列卓有成效的工作。本文介绍语义 Web 的基本概念,阐述语义 Web 在中医药领域的应用现状和发展前景,为该技术在中医药领域的推广工作提供参考。

关键词 语义 Web; 中医药信息学; 中医药; Web 应用

Review on Application of Semantic Web in Traditional Chinese Medicine

Yu Tong, Cui Meng, Li Jinghua

(Institute of Information on Tradition Chinese Medicine, Academy of Chinese Medicine Science, Beijing, Post code: 100700)

Abstract The Semantic Web is an emerging web technology which establishes a global data network to support more intelligent applications. TCM communities have achieved a lot in generalizing the Semantic Web in Traditional Chinese Medicine (TCM) field. This paper summarizes introduction to the Semantic Web, present status and developing perspectives of the semantic web application in TCM field, so as to provide reference to the generalization of this technology in TCM field.

Key Words Semantic Web; TCM informatics; Traditional Chinese medicine; Web application

doi:10.3969/j.issn.1673-7202.2013.01.042

中医药信息化建设是中医药知识传承与创新发展的的重要手段。国家对此高度重视、持续投入,建立了为数众多、内容丰富、规模巨大的中医药科学数据库群,基本实现了中医药知识遗产的系统性保护。但这些资源也存在着独立封闭、零散分布、异质异构等问题,严重地阻碍了知识的跨域流动和持续创新。其中的主要问题包括:1)缺乏适合自身特点的知识建模方法;2)数据资源存在质量缺陷,整合程度较低;3)缺乏有效的知识共享与传播手段,知识资源过分闲置;4)缺乏知识创造与发现的有效手段。语义 Web (Semantic Web, 简称 SW) 为解决这些问题提供了有效的技术手段。

语义 Web 是 World Wide Web Consortium (简称 W3C) 提出并倡导使用的一项扩展性技术。传统 Web 的主要缺陷,是缺乏对结构性数据 (Structured Data) 的支持。为此,Web 创始人蒂姆·伯纳斯-李 (Tim Berners-Lee, Tim BL) 等人于 2001 年在《科学美国人》上正式提出了“Semantic Web (简称 SW)”的构想^[1],其核心思想是在 Web 上构建一个全球性的数据网络,用以实现更为智能的 Web 应用^[1-2]。近年来,SW 技

术在欧、美等地取得迅速发展,得到 IBM、Oracle 等著名 IT 企业的鼎力支持,在生命科学、医疗保健和新药研发等领域取得广泛应用^[3]。受其影响,中医团体开展了一系列 SW 技术的推广工作,初步证实了在中医药领域中使用 SW 技术的必要性和可行性^[4-6]。本文将介绍语义 Web 的基本概念,阐述语义 Web 在中医药领域的应用现状和发展前景,为该技术在中医药领域的推广工作提供参考。

2 语义 Web 在中医药领域发挥的作用

SW 被视为“机器可以理解的 Web”。它是对 Web 架构的系统性扩展,使 Web 支持结构性数据的发表、共享和关联,提高 Web 的有序性,支持智能代理的实现。作为 SW 的基石,领域本体 (Domain Ontology) 是针对某个领域的概念体系的精确规范,用以明确概念的定义以及概念之间的语义关系^[2,7]。Web 本体是通过 Web 在领域团体中共享的本体,为 SW 提供了标准术语,使得 SW 上的数据具有规范的表达方式,便于数据资源的互联、共享与重用。

SW 的核心工作是整合各类数据资源,构建全球数据网络。2006 年,Tim BL 等人发起了“开放关联数据

基金项目:国家重大科技计划“重大新药创制”重点专项——综合性中药新药研究开发技术大平台项目子课题——中药新药发现、新药设计和信息平台 (编号:2009ZX09301-005-01); 国家 863 项目——高效能计算机及网络服务环境——中国中医药科学数据网格服务应用 (编号:2006AA01A122)

通信作者:于彤,中国中医科学院博士后研究人员,隶属于中国中医科学院中医药信息研究所原理与标准研究室,北京市东直门南小街 16 号,邮编:100700, E-mail: yutongoracle@hotmail.com

(Linked Open Data, LOD)”这一大型国际合作项目,号召各界团体、个人广泛参与,通过在线协作的方式,将各种公开数据(它们在数据格式、编码方式和访问方法等方面存在异构性)统一转为 SW 格式并重新发表^[8]。正如关系型数据库支持 SQL 查询,LOD 支持 SPARQL 查询^[8]。SPARQL 可被视为概念层次上的查询,结构简单、易于理解,能有效的屏蔽不同数据模式之间的结构差异,同时又富含语义,能清楚表达用户的查询意图,提供了强大的 LOD 数据访问能力。当前,LOD 在生命科学、医疗保健、新药研发等领域取得成功的应用^[9],成为 SW 领域最具影响力的工程,为实现 SW 的构想迈出了坚实的第一步。

SW 技术在中医药领域具有广阔的应用前景^[4-5]。下面从 3 个方面讨论 SW 技术在中医药信息化建设中发挥的作用。

2.1 中医药知识建模 中医药行业尚缺乏被广泛接受的标准化知识表示模型,这严重地阻碍了知识的交换、共享与传播。中医专家正在使用领域本体作为中医药知识建模的有效手段。例如,尹爱宁等开发了《中医药一体化语言系统》这样一个面向中医药领域的大型领域本体^[10-11],证明了本体工程方法在中医药领域的可行性。

在中医药本体工程实践中,我们发现中医药领域具有复杂性、模糊性和整体性等特点,需要研发与之相适应的特定方法。为此,毛氏、宓氏、于氏等从中医药领域需求出发,对国际最新的 OWL 表达框架做出了一系列的改进和扩展,将 SW 技术成功引入中医药本体工程中。例如,毛氏提出了子本体模型,设计了基于子本体的缓存、演化和推理方法,较好地解决了中医药本体管理与重用问题^[12]。宓氏等基于这些方法构建了中医药本体共建系统 DartOnto,支持 Web 本体的协作式在线加工^[13]。于氏等使用 DartOnto 系统,构建示范性本体,对中医药理论知识进行辨认、梳理、澄清和永久保真处理^[5]。本体建模的对象包括阴阳、五行、脏腑、证候、草药、方剂等基本概念,以及五行学说、藏象学说、辩证论治和方剂配伍等理论学说。

于氏、宓氏等提出基于 SW 的中医药术语融合方案,开发并部署了中医药多术语融合系统,将《中医临床术语集》《中医药学主题词表》和《中医药学名词》等多项术语系统统一转换为 SKOS 格式,实现中医药领域主要术语系统的融合,面向中医团体提供统一的术语服务^[5,13]。

上述实践表明,SW 技术能够提供友好、高效的开发工具,辅助领域专家协同工作,保证大型本体开发的

质量。

2.2 中医药知识服务 中医药知识主要存储于关系型数据库,它们往往服务于特定的医疗和研究机构,无法被其他机构访问,形成“数据孤岛”现象^[14]。

Cheung K. 等提出使用 SW 作为中医药数据表示标准,在 LOD 的基础上实现中西医领域的数据互联,解决“数据孤岛”的问题^[6]。为实施这一策略,Chen H. 等研制了 DartGrid 工具集,并使用 DartGrid 工具构建了大型数据库集成系统,该系统整合 70 多个数据库,内容覆盖中医学、中药学、针灸学等几乎所有中医药门类,成为全球知名的中医药知识融合与共享平台,面向中医团体提供复杂、跨域、普适的知识服务^[15-17]。

付氏等开发了语义搜索引擎 DartSearch,该系统提供丰富的内容和强大的搜索功能,有效满足 Web 用户的知识检索需求^[18]。盛氏等开发了语义查询系统 DartQuery,向 Web 用户提供简易、友好的查询构造界面,引导用户以交互的方式构造 SPARQL 查询^[19]。宓氏等使用语义维基技术开发了中医药百科系统,支持知识资源的有序组织、有效管理和协作式加工^[13]。

2.3 中医药知识发现 通过知识发现方法,能从海量中医药数据中发现新颖知识(模式与规则等),以辅助新药研发和临床决策等活动^[20]。SW 能更好地解决中医药领域的数据集成问题,为中医药知识发现提供了丰富的数据资源^[21]。吴氏等基于语义网格技术框架,研制了中医药知识发现平台,支持面向海量数据的数据挖掘与知识发现^[14]。该系统在利用分子生物学研究中医药物成份和功能方面取得了具有启发性的结果。

中医药领域涉及大量实体之间的复杂关系,如证候和基因之间的对应关系、方剂与中药之间的组成关系,中药之间的相互作用关系等。这些关系适合被表示为具有复杂拓扑结构的网络^[22]。陈氏、于氏等提出了语义图挖掘方法论,该方法论使用本体和语义图作为领域知识表示的主要手段,结合图挖掘和机器推理等方法来分析复杂网络,并挖掘有意义的模式和规则^[23-24]。于氏等在面向 LOD 的分布式计算环境中实现了这一方法论,用于分析中药相互作用网络、证候-基因网络等中医药复杂网络^[21],初步验证了在中医药知识发现领域使用 SW 技术的可行性。

3 语义 Web 在中医药领域的应用案例

3.1 中医药语义电子科学环境(TCM Semantic e-Science Environment, TCM-SESE) TCM-SESE 是一个大型的电子科学平台,该系统实现了中医药知识资源的有效集成和综合管理,面向全国 40 多家机构提供中

医药知识服务^[17]。该工程首次将 SW 技术系统引入中医药领域,具体工作包括:1)采用 Web 本体技术,构建中医药领域本体,解决中医药知识建模问题;2)采用语义映射和查询重写等技术,实现大量中医药数据库的有效整合,解决中医药“数据孤岛”问题;3)采用语义查询、语义搜索、语义维基等技术,面向临床决策、新药研发和电子教学提供知识服务,解决中医药知识传播问题。TCM-SESE 现已投入使用并稳定运行,实现中医药知识资源的深度整合和充分共享,促进中西医领域专家之间的交流与合作,在临床决策、新药研发和电子教学等案例中发挥了重要作用。

3.2 Linking Open Drug Data (LODD)^[24] LODD 是一个大型国际合作项目,旨在利用 LOD 的技术框架,整合医药学数据资源,加速新药研发的进程。针对传统中草药的研究,可能对现代新药研发提供重要线索,但这些宝贵研究结果往往缺乏整合和应用^[25]。为此,LODD 的一项重要任务是将中医药知识发布在 LOD 上,并建立中西医之间的知识关联,支持中药新药研发等应用^[26-27]。LODD 中记录了草药的知识(如化学成分、疗效等),以及疾病、基因、草药之间的关联关系,能辅助研究人员找出能够治疗某种疾病的草药,理解草药的性质和疗效,支持中药新药研发等应用。LODD 被成功用于“研究具有抗抑郁功效的中草药”这一实际案例:通过 SPARQL 查询从 LOD 中查找出可能治疗抑郁的中草药,检索它们的活性成分,作为新药研发过程的先导化合物^[28]。

这两个项目在海量中医药数据处理方面积累了丰富的经验,在中西医之间建立了一道交流的桥梁,验证了 LOD 技术方案的可行性以及 SPARQL 查询语言和协议的灵活性,为后续工作起到了良好的示范作用。

4 结束语

SW 正在成为一项主流技术,获得广泛的认可和采用。中医团体已开展了 SW 的推广工作,如中医药语义电子科学环境、LODD 等,为保护中医药文化遗产,促进中西医结合,推动中医药现代化做出重要贡献,展示了 SW 在跨组织、跨领域、跨文化应用中的技术优势。展望未来,我们相信 SW 技术将在跨文化、多学科交叉的研究中发挥更为重要的作用。

参考文献

- [1] BERNERS-LEE T, HENDLER J, LASSILA O. The Semantic Web [J]. Scientific American, 2001, May.
- [2] ALLEMANG D, HENDLER J. Semantic Web for the Working Ontologist [M]. Elsevier, 2011.
- [3] DOMINGUE J, FENSEL D, HENDLER J. Handbook of Semantic Web Technologies [M]. Springer, 2011.
- [4] 吴朝晖. 语义 Web 与中医药 [N]. 计算机世界, 2007-11-26.
- [5] 于彤. 知识服务:语义 Web 在中医药领域的应用研究 [D]. 浙江大学, 2012.
- [6] CHEUNG K, CHEN H. Semantic Web for data harmonization in Chinese medicine [J]. Chinese Medicine 2010, 12(5):2.
- [7] HORROCKS I. Ontologies and the semantic web [J]. Commun. ACM, 2008, 51(12): 58-67.
- [8] HEATH T, BIZER C. Linked Data: Evolving the Web into a Global Data Space [M]. Morgan & Claypool, 2011.
- [9] BIZER C, HEATH T, BERNERS-LEE T. Linked data - the story so far [J]. Int. J. Semantic Web Inf. Syst., 2009, 5(3):1-22.
- [10] 尹爱宇, 张汝恩. 建立<中医药一体化语言系统> [J]. 中国中医药信息杂志, 2003, 10(3):90-91.
- [11] ZHOU X, WU Z, YIN A, WU L, FAN W, ZHANG R. Ontology development for unified traditional Chinese medical language system [J]. Artif Intell Med, 2004, 32(1):15-27.
- [12] MAO Y, WU Z, TIAN W, et al. Dynamic sub-ontology evolution for traditional Chinese medicine web ontology [J]. Journal of Biomedical Informatics, 2008, 41(5):790-805.
- [13] 苻金华. 中医药知识工程应用 [D]. 浙江大学计算机科学与技术学院, 2010.
- [14] 吴朝晖, 陈华钧. 语义网格:模型、方法与应用 [M]. 杭州:浙江大学出版社, 2008.
- [15] CHEN H, WANG Y, WANG H, et al. From Legacy Relational Databases to the Semantic Web: an In-Use Application for Traditional Chinese Medicine [C]//International Semantic Web Conference, 2006.
- [16] CHEN H, WU Z, MAO Y, ZHENG C. DartGrid: a semantic infrastructure for building database Grid applications [J]. Concurrency and Computation Practice and Experience, 2006, 18(14): 1811-1828.
- [17] CHEN H, MAO Y, ZHEN X, et al. Towards semantic e-science for traditional Chinese medicine. BMC Bioinformatics, 2007, 8(Suppl 3): 56.
- [18] FU Z, CHEN H, YU T. Intelligent search on integrated knowledge base of traditional Chinese medicine. Journal of Southeast University (English Edition) [J]. 2009, 25(4): 460-463.
- [19] 盛浩. 基于超链数据的中医药语义查询系统 [D]. 浙江大学, 2011.
- [20] FENG Y, WU Z, ZHOU X, ZHOU Z, FAN W. Knowledge discovery in traditional Chinese medicine: State of the art and perspectives. Artificial Intelligence in Medicine, 2006, 38(3):219-236.
- [21] WU Z, YU T, CHEN H, et al. Semantic Web Development for Traditional Chinese Medicine [C]//AAAI-08/LAAI-08, 2008, 1757-1762.
- [22] ZHOU X, LIU B, WU Z, FENG Y. Integrative mining of traditional Chinese medicine literature and MEDLINE for functional gene networks [J]. Artificial Intelligence in Medicine, 2007, 41, 87-104.
- [23] CHEN H, DING L, WU Z, TONG Y, et al. Semantic web for integrated network analysis in biomedicine [J]. Briefings in Bioinformatics, 2009, 10(2):177-192.
- [24] YU T, JIANG X, FENG Y. Semantic Graph Mining for e-Science [C]//AAAI workshop on Semantic e-Science, 2007, 77-80.
- [25] FANG Y, HUANG H, CHEN H, JUAN H. TCMGeneDIT: a database for associated traditional Chinese medicine, gene and disease information using text mining [J]. BMC Complement Altern Med, 2008, 8:58.
- [26] ZHAO J, JENTZSCH A, SAMWALD M, CHEUNG K. Linked data for connecting traditional Chinese medicine and Western medicine [C]. The Sixth International Workshop of Data Integration in the Life Sciences (Poster&Demo). Manchester, UK 2009, 13.
- [27] ZHAO J. Publishing Chinese medicine knowledge as Linked Data on the Web [J]. Chinese Medicine 2010, 5(1):27.
- [28] SAMWALD M, DUMONTIER M, ZHAO J, et al. Integrating findings of traditional medicine with modern pharmaceutical research: the potential role of linked open data [J]. Chinese Medicine, 2010, Dec17(5): 43.

(2012-06-27 收稿)