

• 医学信息研究 •

中医药知识图谱构建*

贾李蓉 刘 静 于 彤 董 燕 朱 玲 高 博 刘丽红

(中国中医科学院中医药信息研究所 北京 100700)

〔摘要〕 在对知识图谱进行简要介绍的基础上,从数据来源、研究内容、图形化展示几方面探讨如何构建中医药知识图谱,实现中医药知识资源的有效整合,最后提出中医药知识图谱的应用前景。

〔关键词〕 知识图谱;中医药;知识服务

〔中图分类号〕 R-058 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2015.08.012

Construction of Traditional Chinese Medicine Knowledge Graph JIA Li-rong, LIU Jing, YU Tong, DONG Yan, ZHU Ling, GAO Bo, LIU Li-hong, Information Institute of Traditional Chinese Medicine, China Academy of Chinese Medical Sciences, Beijing 100700, China

〔Abstract〕 Based on introducing knowledge graph, the paper discusses how to construct Traditional Chinese Medicine (TCM) knowledge graph from the aspects of data sources, research contents and visualized display, so as to integrate TCM knowledge resources. It also puts forward the application prospect.

〔Keywords〕 Knowledge graph; Traditional Chinese Medicine (TCM); Knowledge service

1 引言

随着知识的爆炸式增长,计算机存储和处理能力的不断提升,信息的存储、分析、检索、可视化等技术的发展,一种以知识单元为基础,能有效获取知识、快速把握学科前沿领域的方法——知识图谱(Knowledge Graph)悄然兴起。知识图谱可以绘制、挖掘、分析和显示科学技术知识以及它们之间相互关系,是在大数据时代背景下产生的一种新型

的海量知识管理与服务模式,涉及应用数学、信息科学及计算机科学诸学科,是科学计量学和信息计量学的新发展^[1]。其研究目标是借助现代技术与理论使知识可视化,让人们更加方便、准确地获取知识。知识图谱作为知识的载体,能用图形化的方式将人们不易理解的信息形象地表示出来^[2],通过内容分析、引文分析和可视化的方式显示知识结构及其相互关系^[3],既符合人类的认知习惯,又充分利用了现代信息技术;使用户既能快速获取知识及其之间的逻辑关系,又能从海量文献中把握关键的知识^[4]。近年来越来越多的中医药数据库、电子资源等都可以在互联网上方便获取,如何获取更准确、更全面、更权威的中医药知识也成为中医药领域的研究需求。本文对中医药知识图谱构建进行示范性研究,以期能为中医药领域知识服务方面提供借鉴。

〔修回日期〕 2015-04-02

〔作者简介〕 贾李蓉,助理研究员,发表论文 20 余篇;通讯作者:刘丽红,助理研究员。

〔基金项目〕 中国中医科学院基本科研业务费自主选题项目“基于中医药学语言系统的知识服务平台研究”(项目编号:ZZ080302)。

2 中医药知识图谱构建

2.1 概述

知识图谱是以语义网络为骨架构建起来的巨型、网络化的知识系统，能捕捉并呈现领域概念之间的语义关系，使互联网上琐碎、零散的知识相互连接，支持综合性知识检索以及问答、决策支持等智能应用。中医药知识图谱主要是以中医药学语言系统为骨架构建的知识图谱体系，以现有的数据库资源为知识图谱填充内容。可视化的语义图可以形象地表达领域概念之间的关联，用户可通过交互的方式来浏览领域概念，选择其中的某个概念开始构造查询或搜索。中医药知识图谱能增强中医药知识资源的联通性，支持中医用户在概念层次上浏览领域知识资源，发现中医药概念或知识资源之间的潜在联系。

2.2 数据来源

知识图谱的数据包括两方面：一方面是知识图谱的骨架，另一方面是整个图谱所涉及的知识。中医药知识图谱的骨架采用中国中医科学院中医药信息所研制的中医药学语言系统。该系统自 2002 年起研制，目前已发展为包含 13 万多个概念、30 余万术语以及 100 余万语义关系的大型语义网络，基本覆盖了中医药学科的概念体系，在规模和完整性等方面都处于中医药界领先地位^[5]。示范研究选取 2010 版药典的药材和饮片部分 611 味中药作为基本数据，然后选用《临证用药配伍指南》、新版中国中药数据库、《有毒中药》、《中药材正名词典》、《方剂大辞典》和临床疾病诊疗数据库中涉及这些中药的数据为知识图谱的数据基础。

2.3 研究内容^[6-7]

(1) 研究基于知识图谱的中医药领域知识组织方法，以中医药学语言系统为骨架，将中医药领域主要的词表、术语资源和单表型数据库的内容导入中医药知识图谱之中，研发海量中医药知识图谱数据处理技术，实现大型中医药知识图谱的存储、查

询和更新。(2) 研发基于中医药语言系统 (Traditional Chinese Medicine Language System, TCMLS) 的中医药文献知识获取技术，研究从海量中医药文献中发现新颖的实体和语义关系的方法，构建交互式中医药文献知识获取工具，以半自动的方式扩展中医药知识图谱的内容。(3) 设计并实现符合中医药领域实际需求的知识图谱检索和浏览方法。通过交互性语义图界面对知识图谱的内容进行展示，协助网络用户浏览概念信息及概念之间的关系。另拟将知识图谱嵌入搜索和维基等系统之中提供知识服务。

2.4 图形化展示

图形化展示是知识图谱的重要部分。本文构建的中医药知识图谱提供了图形化的展示，主要对中药概念进行语义网络、图片和相关知识的展示等。知识图谱界面最上部分是搜索框，如搜索“人参”，系统界面，见图 1。

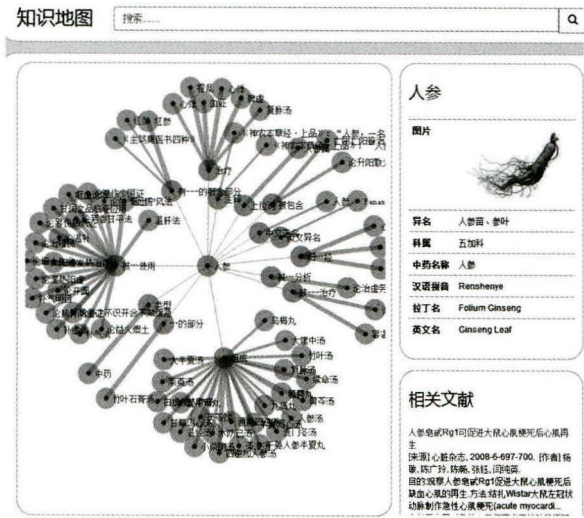


图 1 中医药知识图谱展示界面

左侧显示以“人参”为中心的相关知识图，该图采用开源的可视化 js 组件库 d3.js，并且使用的是 svg 矢量图，可以无损缩放，兼容各种分辨率。右侧上半部分为“人参”的基本信息，点击打开可显示百科知识，见图 2。右侧下半部分则提供了相关的文献检索结果，该结果目前为“中国中医药期刊文献数据库”中的全文检索结果。



图 2 百科知识示例

3 应用前景

3.1 基于中医药知识图谱的检索系统^[7-8]

中医药知识图谱增加了中医药领域知识资源的连接性，支持中医药用户在概念层上浏览领域知识资源，发现中医药概念知识间的潜在联系。可基于中医药知识图谱在传统的检索系统中嵌入“知识图谱卡片”，系统会根据用户的输入识别出概念，列出概念的核心信息和相关信息。用户可根据不同需求了解概念及其相关知识，在相关概念之间进行跳转，从而对检索点进行全方面的了解。这种基于知识图谱的检索系统不仅直接提供给用户输入的知识，而且提供了相关的信息，方便检索。

3.2 基于中医药知识图谱的知识地图^[9]

可基于中医药知识图谱实现中医药知识地图系统，知识地图以可视化概念图的方式形象地表达领域概念之间的关联，用户可通过交互式的方式来浏览中医药相关概念，选择感兴趣的概念查询浏览。知识地图能够实现对中医药知识的有效组织和导航，协助用户快速找到所关注的知识点，有可能帮助用户发现潜在的知识等。

3.3 基于中医药知识图谱的维基百科^[10]

可基于知识图谱搭建中医药维基百科系统，面向中医药工作者和用户提供百科全书式的知识服务。中医药维基百科系统按概念实体对知识和文献进行组织，将知识图谱中关于某个概念实体的知识以百科的形式综合呈现出来，包括该概念的名称、类型、简介、文字信息以及语义关系等，列出与该实体相关的文献题录。相较于传统的维基百科，基于中医药知识图谱的维基百科是利用知识图谱将已有的中医药信息资源有序组织起来，提供知识点的关联性，具有自动提供知识并能追溯知识源的特点。

4 结语

中医药知识图谱是以中医药学语言系统为骨架、以现有数据库资源为填充内容构建的知识图谱体系。本文探讨了中医药知识图谱构建的数据来源、研究内容及应用前景等内容，但鉴于中医药知识图谱的构建研究尚为探索阶段，还有很多不足之处，今后需从以下几个方面进行改进：（1）继续完善和发展中医药学语言系统。以本研究的方法来构建知识图谱，中医药学语言系统作为系统的“骨架”是最重要的元素，继续完善和发展中医药学语言系统是本工作的基础。（2）进行数据资源的映射规则和数据元等标准的研究。由于知识图谱要将众多不同格式、不同数据库的资源整合，故而多种数据资源如何映射以及数据库的数据元等标准的

（下转第 59 页）

- 2 中国自闭症儿童发展状况报告 [EB/OL]. [2014-12-27]. <http://www.guduzheng.net/2014/10/38086448.html>.
- 3 宋立荣, 齐娜, 张群. 网络健康信息传播的信息质量问题思考 [J]. 医学信息学杂志, 2014, 35 (10): 8-12.
- 4 张士靖, 杜建. 健康信息素养应成为中国公众健康素养促进的关键点 [J]. 医学信息学杂志, 2010, 31 (2): 45-48.
- 5 American Psychiatric Association. Diagnostic and statistical manual of mental disorder [M]. Washington DC: American Psychiatric Publishing, 2013.
- 6 Kanner L. Autistic Disturbances of Affective Contact [J]. Nervous Child, 1943, 38 (2): 217-250.
- 7 王安莲, 刘志荣. 自闭症研究现状 [J]. 安徽预防医学杂志, 2013, 19 (5): 367-382.
- 8 杨黄芬. 自闭症学生之教育 [M]. 台北: 心理出版社, 2005.
- 9 Jorde L B, Hasstedt S J, Ritvo E R, et al. Complex Segregation Analysis of Autism [J]. Am J Hum Genet, 1991, (49): 932-938.
- 10 Davidovitch M, Glick L, Holtzman G, et al. Developmental Regression in Autism: maternal perception [J]. J Autism Dev Disord, 2000, (30): 113-119.
- 11 沈丽宁. 国外健康信息服务现状扫描及启示 [J]. 医学信息学杂志, 2010, 31 (6): 38-51.
- 12 彭琰, 严莉, 朱红. 大数据时代用户健康信息学的价值 [J]. 医学信息学杂志, 2014, 31 (1): 2-6.
- 13 张会会, 马敬东, 邱金平. 网络健康信息质量评估研究综述 [J]. 医学信息学杂志, 2014, 35 (4): 2-5.
- 14 PubMed Health [EB/OL]. [2014-12-02]. <http://www.ncbi.nlm.nih.gov/pubmedhealth/PMH-0002494/>.
- 15 US Autism&Asperger Association [EB/OL]. [2014-12-02]. <http://www.usautism.org/tv/>.
- 16 Gazan R. Social Q&A [J]. Journal of the American Society for Information Science and Technology, 2011, 62 (12): 2301-2312.
- 17 Kim S, Oh S. Users' Relevance Criteria for Evaluating Answers in a Social Q&A Site [J]. Journal of the American Society for Information Science and Technology, 2009, 60 (4): 716-727.

(上接第 53 页)

研究也就成了知识图谱构建的重要环节。(3) 知识图谱系统的功能还有待继续完善。目前对数据的操作进行了很多后期人工的规范, 需要深入到数据库中修改数据, 以后将完善数据管理等系统功能。

参考文献

- 1 刘则渊, 陈悦, 侯海燕, 等. 科学知识图谱: 方法与应用 [M]. 北京: 人民出版社, 2008: 16-26.
- 2 陈悦, 刘泽渊, 陈劲, 等. 科学知识图谱的发展历程 [J]. 科学学研究, 2008, (3): 449-460.
- 3 孙小欣. 基于潜在语义分析的学科知识图谱构建 [D]. 武汉: 华中师范大学, 2013.
- 4 唐钦能, 高峰, 于金平. 知识地图相关概念辨析及其研究进展 [J]. 情报理论与实践, 2011, (1): 121-125.
- 5 贾李蓉, 于彤, 崔蒙, 等. 中医药学语言系统研究进展 [J]. 中国数字医学, 2014, 9 (10): 57-59, 62.
- 6 冀玉静, 李军莲, 李芳. 词共现分析在构建概念空间中的应用研究综述 [J]. 医学信息学杂志, 2010, 31 (10): 49-52.
- 7 董燕, 李海燕, 崔蒙, 等. 中医临床术语系统建设概况与改进措施 [J]. 医学信息学杂志, 2014, 35 (8): 43-48.
- 8 刘俊丽, 张秀梅, 蒋勇青. 基于文本挖掘的乙型肝炎相关文献知识图谱分析 [J]. 医学信息学杂志, 2014, 35 (1): 48-53.
- 9 王敏, 张燕舞, 张玢. 信息可视化在医学文献分析中的初步应用理论研究 [J]. 医学信息学杂志, 2010, 31 (2): 40-49.
- 10 陈兰杰, 董芳. 基于知识图谱的国际竞争情报研究热点与前言的信息可视化分析 [J]. 医学信息学杂志, 2010, 31 (8): 7-11.