doi:10.3772/j.issn.2095-915x.2017.01.004

基于本体的中医知识图谱构建

1. 北京科技大学计算机与通信工程学院　北京　100083；
2. 材料领域知识工程北京市重点实验室　北京　100083
张德政[1,2]　谢永红[1,2]　李曼[1,2]　石川[1,2]

摘 要　为了更加有效地分析中医药知识之间的联系，优化知识的检索，共享中医领域知识，使中医更好的进行传承。本文提出了基于本体的中医核心知识图谱表示及其构建方法，研究了中医本体与知识图谱的映射方法，实现了于中医本体的中医核心知识图谱的构建，并进一步研究和实现多源知识获取技术及基于知识图谱的名老中医临证经验的发现，为进一步构建中医领域全面的知识图谱，挖掘整理中医临证经验与学术思想及建立基于信息检索技术的中医知识服务打下坚实基础。

关键词：中医，本体，知识图谱
中图分类号：TP182

Construction of Knowledge Graph of Traditional Chinese Medicine Based on the Ontology

1. School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China;
2. Beijing Key Laboratory of Knowledge Engineering for Materials Science, Beijing 100083, China
ZHANG DeZheng[1,2]　XIE YongHong[1,2]　LI Man[1,2]　SHI Chuan[1,2]

Abstract　In order to analyze more effectively the connection between the Traditional Chinese Medicine (TCM) knowledge, to optimize the knowledge retrieval, and to share the domain knowledge of TCM, so as to better spread and inherit the TCM, this article put forward representation and construction method of knowledge graph of TCM based on the ontology. We also studied the mapping method between the ontology and knowledge map, achieved the construction of the core of knowledge graph of TCM based on the traditional Chinese Medicine ontology, and further studied and realized the multi-source knowledge

基金项目：本文受国家科技支撑计划项目：名老中医临床经验、学术思想传承研究（一）（2013BAI13B06）的资助。
作者简介：张德政（1964–），教授，研究方向：大数据与知识工程；谢永红（1970–），通讯作者，副教授，研究方向：知识发现与知识工程，email：xieyh@ustb.edu.cn；李曼（1993–），硕士研究生，研究方向：知识工程；石川（1992–），硕士研究生，研究方向：知识工程。

doi:10.3772/j.issn.2095-915x.2017.01.004

acquisition technology and the discovery of famous traditional Chinese Medicine clinical experience based on the knowledge graph. Moreover, this study also lay a solid foundation for further constructing the comprehensive knowledge graph of TCM, excavating and tidying the traditional Chinese Medicine clinical experience and academic thought and establish traditional Chinese Medicine knowledge service based on the technology of information retrieval.

Keywords：Traditional Chinese Medicine, ontology, knowledge graph

1 引言

中医历经了几千年的发展过程，积累了大量珍贵的临床经验，形成了无数经典理论。中医药知识传承的前提条件，是对中医基础理论和知识体系的整理和分析。如何借助信息科学与技术来对中医药理论和知识体系进行整理和分析，对隐含在医案和文献中的学术思想、临床经验和辨证方法进行挖掘，是值得中医学者探索的一个重要问题。随着计算机科学技术的不断发展，知识工程领域引进了知识图谱概念。知识图谱能够描述现实世界的实体及其之间的关系，能够实现对知识的共建、共享及重用。针对知识图谱这一特点，引用知识图谱解决中医学在知识表达、共享和应用方面的问题也受到中医界的重视。于彤等提出了以 TCMLS 为骨架，以中医药领域现有的术语和数据库资源为内容，构成大型知识图谱的构想，并开展了相关的探索和实践，但是没有实现中医药知识资源的有效整合以及提供全面、及时、可靠的知识服务[1]。阮彤等利用文本抽取、关系数据转换以及数据融合等技术探索了中医药知识图谱自动化构建的方法[2]。贾李蓉等从数据来源、研究内容、图形化展示探讨了如何构建中医知识图谱，但是如何从多个数据来源构建知图谱没有给出具体描述，相关应用也停留在浏览检索方面，没有进行数据资源的映射规则和数据元等标准的研究[3]。

本文基于前期建立的中医基础理论本体，结合知识图谱技术，结合图数据库特点，提出并实现中医核心知识图谱表示和构建技术，在基于图谱在结合中医思维方法的知识检索和名老中医经验发现和总结方面进行有益的探索。

2 相关知识及研究基础

2.1 本体

本体作为一种知识表示方法，与谓词逻辑、框架（frame）等其他方法的区别在于他们属于不同层次的知识表示方法，本体表达了概念的结构、概念之间的关系等领域中实体的固有特征，即“共享概念化”，而其他的知识表示方法如语义网络等，可以表达某个体对领域中实体的认识，不一定是实体的固有特征。这正是本体层与其他层次的知识表示方法的本质区别[4,5]。

本体的每一个知识表示元素也可以被看作一个知识片，每一个知识片都包含名称、定义和文档说明[6]。将本体引入知识库的知识建模[7]，建立领域本体知识库，可以用概念对知识进行表示，同时揭示这些知识之间内在的关系。领域本体知识库中的知识，不仅通过纵向类属分类，而且通过本体的语义关联进行组织和关联，

推理机再利用这些知识进行推理，从而提高检索的查全率和查准率。

2.2 知识图谱

近几年来，随着 Linking Open Data 等项目的全面展开，语义 Web 数据源的数量激增，大量 RDF 数据被发布[8, 9, 10]。互联网正从仅包含网页和网页之间超链接的文档万维网(Document Web ）转变成包含大量描述各种实体和实体之间丰富关系的数据万维网（ Data Web ）。在这个背景下，于 2012 年 5 月 Google 正式提出知识图谱（ Knowledge Graph ），并建立以知识图谱为基础的新搜索功能，用以改善搜索结果，提供更加快捷的搜索体验。自此，它在学术界和工业界掀起了一股热潮。国内外各大互联网企业在之后的短短一年内纷纷推出了自己的知识图谱产品以作为回应。例如百度的“知心”，搜狗的“知立方”等搜索引擎公司用来改进搜索质量，从而拉开了语义搜索的序幕。

知识图谱是描述真实世界中存在的各种概念或实体，以及各种实体、概念之间的关系。其中，每个概念或实体用一个全局唯一确定的 ID 来标识，称为它们的标识符（ identifier ），这种做法与一个网页有一个对应的 URL、数据库中的一条记录有一个特定的主键相似；实体可以拥有属性，用于刻画实体的内在特性，每个属性都是以“〈属性，值〉对（ Attribute-Value Pair，AVP ）”的方式来表示；而关系（ relation ）用来连接两个实体，刻画它们之间的关联。

知识图谱本质上是语义网络，是一种基于图的数据结构，由节点（ Point ）和边（ Edge ）组成，即知识图谱亦可被看作是一张巨大的图，在知识图谱里，每个节点表示真实世界中存在的概念或实体，每条边则表示属性或实体之间的关系。知识图谱是关系的最有效的表示方式。通俗地讲，知识图谱就是把所有不同种类的信息（ Heterogeneous Information ）连接在一起而得到的一个关系网络。知识图谱提供了从“关系”的角度去分析问题的能力。

2.3 知识图谱与本体的关系

知识图谱在本体的基础上进行了丰富和扩充，扩充主要体现在实体（ Entity ）层面；本体中突出和强调的是概念以及概念之间的关联关系，而知识图谱则是在本体的基础上，增加了更加丰富的关于实体的信息。模式（ Schema ）是对知识的提炼，而且遵循预先给定的模式有助于知识的标准化，更利于查询等后续处理。本体描述了知识图谱的数据模式（ Schema ），即为知识图谱构建数据模式相当于为其建立本体。

本体是概念层次上面的表示，即侧重表示概念与概念之间的关系；而知识图谱是以实体为核心，事实用实体之间的关系表示，复杂关系可以体现在关系之间组合聚合等关系上，即注重体现实体本身之间的关系推理。

本体[11]能较好的对知识表示进行概括性、抽象性的描述。而知识图谱则能更好的融合关系表示和语义网表示进而组成知识的可拓网络。

3 中医基础理论本体

本文是在课题组前期建立的中医基础理论本体基础上进行研究的，该本体是在中医专家

doi:10.3772/j.issn.2095-915x.2017.01.004

的参与指导下，以普通高等教育中医药类规划教材为主要知识源，包括《中医诊断学》、《中医基础理论》、《方剂学》、《中药学》等，并结合中医临床诊疗术语标准等，使用本体构建工具 Protégé[12] 进行构建的。

该本体以阴阳五行学说为指导，以五脏为中心，包括了中医认识方法、中医生理、中医病理、辨证论治四大部分。这四大部分是一个有机的整体，关系密切。其中，中医认识方法本体概念包括阴阳、五行概念，贯穿指导其他三大部分；中医生理本体概念指人体组成部分的相关概念，包括五脏、六腑、奇恒之腑、气、血、津液、精、体液、外荣、形体、官窍、情志、神、经络穴位等多层级概念；中医病理本体概念包括疾病、病因、病机、症状等多层级子概念，是中医生理的异常情况。辨证论治包括辨证方法、证候、治则、治法、方药、性味归经等概念，与中医病理之间是治疗关系。

中医基础理论知识本体的语义关系结构如下：

1）概念之间通用语义关系

本体中的语义关系表示概念与概念之间的关联关系。本体概念之间的通用关系如表 1 所示。

表1 通用语义关系表

通用关系名	语义关系描述
Kind-of	概念术语之间的继承机制，也可以理解成包含与被包含的关系。
Part-of	概念术语之间存在整体与部分的关系，即子概念描述的事物是父概念描述事物的一部分。如，“辨证方法”包含“八纲辨证”。
Instance-of	概念术语与实例之间的关系。如，“心”是“五脏”的一个实例。
Attribute-of	概念术语与属性之间的关系。如，“药味”是“药”的一个属性。

2）概念之间自定义语义关系

在中医基础理论本体中，根据本体语义关系的层级和结构，得到语义关系主要包括以下五大类，分别是：功能、时间、概念、物理和空间相关性。

自定义语义关系如表 2 所示。

4 中医知识图谱构建

4.1 中医知识图谱表示方法

整个中医的概念体系中，类关系、整体与部分的关系是概念体系的主要关系，所以中医知识图谱的结构分为了两大类关系，一类是概念层级关系、一类是实体关系。

中医知识图谱的基本结构由概念层次关系图 GM 与实体关系图 GE 组成，即 KG=<GM，GE>。

其中，概念层次关系图表示中医本体概念层级结构；实体关系图表示中医实体及其之间的关系。

概念层次关系图 GM=<CM，RM>，其中 CM 表示图中概念节点，RM 表示由多条边连接的两个概念之间的关系边。

实体关系图 GE=<EE，RE>，其中 EE 表示图中实体节点，RM 表示由多条边连接的两个实体之间的关系边。

表2 自定义语义关系表

	自定义语义关系	语义关系注释
物理	方药关系、病症关系、证症关系	组成 / 由组成的关系
	藏	包含 / 被包含的关系
	络属关系	连接 / 相互连接
空间	位于、病位关系	位置
	主、司、运化、行、摄、统、纳	管理 / 被管理的关系
	药证关系、药症关系、药病关系、方证关系、方症关系、方病关系	治疗 / 被治疗的关系
	伤、困、易伤、犯、袭、闭	干扰 / 被干扰的关系
	并发	并发
	相侮、相乘、相克、相生、通应	相互作用
功能	表里关系	互为表里
	预防关系	提前防止
	生	产生 / 被产生的关系
	因病关系、因证关系、因症关系、机证关系、发、母病及子、子病及母	导致 / 引起 / 被引起的关系
	传变	进程
	多、开窍于、华在、在液为、在体为、在志为	现象表达
	治愈、转归	结果 / 效果
时间	兼夹	同时发生 / 先于发生 / 后于发生
	过、虚、实	程度 / 等级
	顺、逆	影响
	诊断	诊断 / 被诊断
语义	属、脏腑生理特性、脏腑生理功能属、脏腑生理特性、脏腑生理功能	特性
	归经关系	归经
	隶属于、包含	概念的隶属包含关系
	机证关系	类似关系（病机与证候）

中医知识图谱的本质是通过概念或实体及其语义关系来表达中医知识的一张巨大的图。中医知识图谱的图结构由节点集合和边集合构成，即 KG = {< N >, <R >}。其中，节点代表表示中医领域知识中的各种概念及实体；边代表概念、实体间的关联，用来连接两个概念或实体。

< N > 表示节点集合，即 N ∈ (CM ∪ EE)。

< R > 表示边集合，即 R = {< T >，< D >，< G >}，R ∈ (RM ∪ RE)。

< T > 表示关系类型集合，< D > 表示关系方向集合，< G > 表示三元组集合，使用三元组（node_A，relation，node_B）表达语义关系，node_A 与 node_B 表示节点（概念或实体），方向是有 node_A 指向 node_B，relation 表示语义关系每个三元组表示一个事实。如图 1 所示虚线部分，心开窍于舌，其中边“开窍于”表示语义关系，其实体是“心”与“舌”。

4.2 本体与知识图谱映射匹配机制

把本体概念层级结构当作树，把本体概念层级结构的组成元素（概念、实例等）作为树的节点，元素之间的继承关系用连线表示。

doi:10.3772/j.issn.2095-915x.2017.01.004

把知识图谱的概念层次关系图当作树，概念节点作为树的节点，节点之间的层级关系用连线表示。把知识图谱的实体关系图当作图，实体节点作为图的节点，实体节点之间的语义关系连线表示。

本体与知识图谱的映射匹配机制则看成是树与树之间的映射，树与图之间的映射。如图2所示。

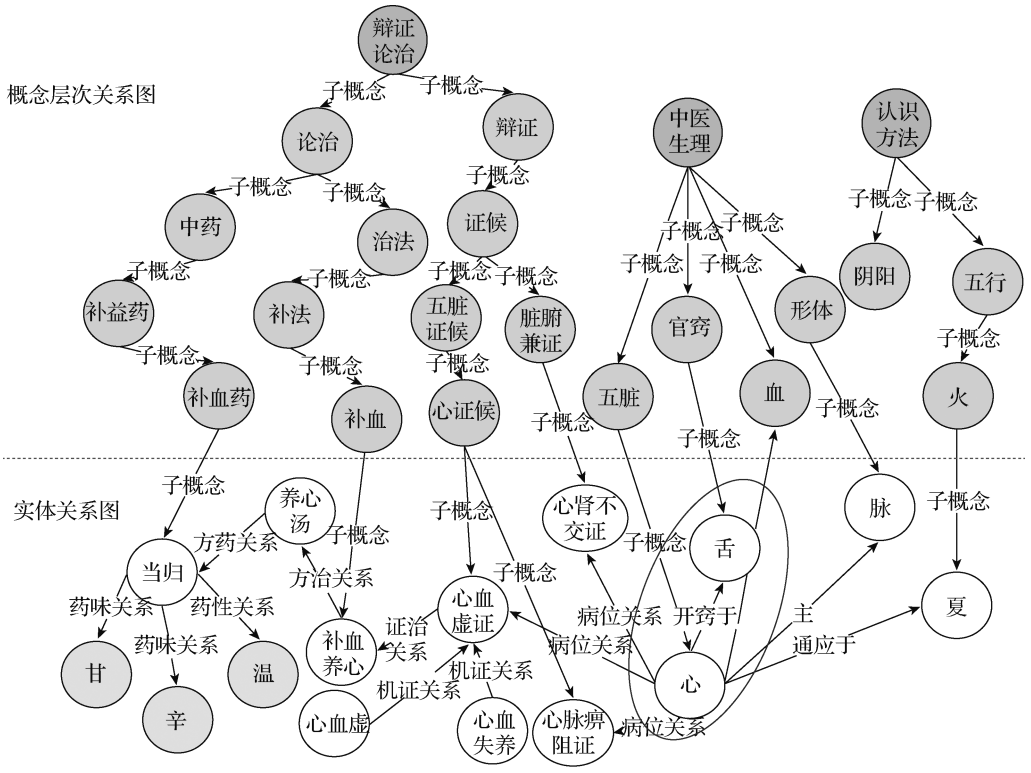

图1 中医知识图谱部分数据图

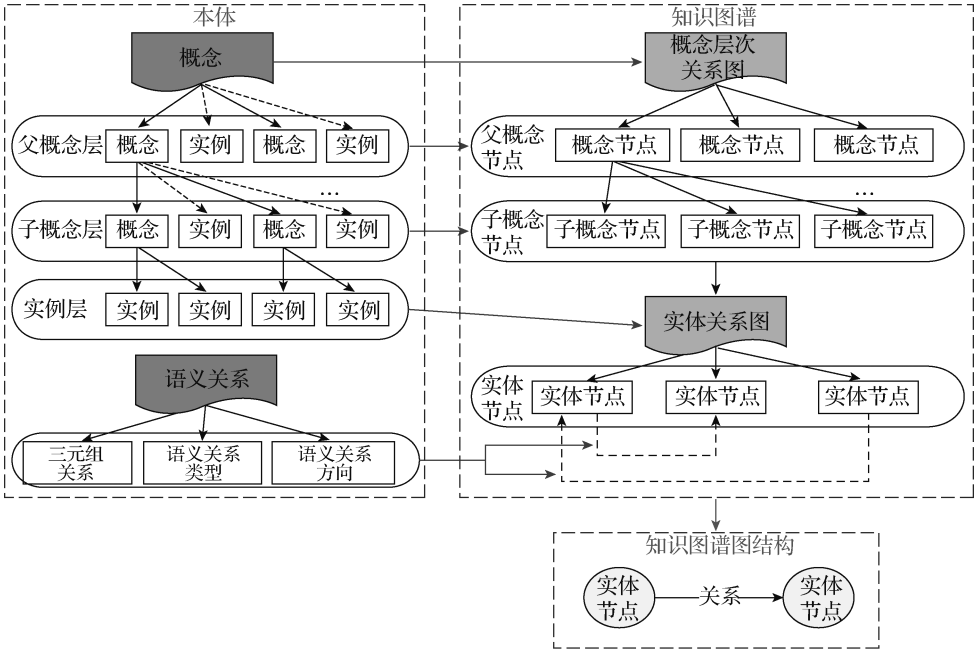

图2 本体与知识图谱的映射匹配机制图

树与树之间的匹配主要考虑树的形状映射。

本体中概念层级结构由顶层概念，多层子概念构成。将本体看做树，则顶层概念为根节点，多层子概念为孩子节点，概念最底层的实例为叶节点，将概念树结构完整的映射到知识图谱中，形成知识图谱的概念层次关系图。

树与图之间的匹配主要考虑节点的位置和路径映射。

本体的实例囊括在概念树结构中，是概念树的叶节点，但其中还具有多种语义关系，将实例表示为图的节点，关系表示为图的边，将其实例及其语义关系完整的映射到知识图谱中，形成知识图谱的基本实体关系图。

4.3 基于本体的中医知识图谱构建方法

从根据以上的中医本体与知识图谱的映射匹配机制和知识图谱的表示，中医核心知识图谱的构建过程如图 3 所示：

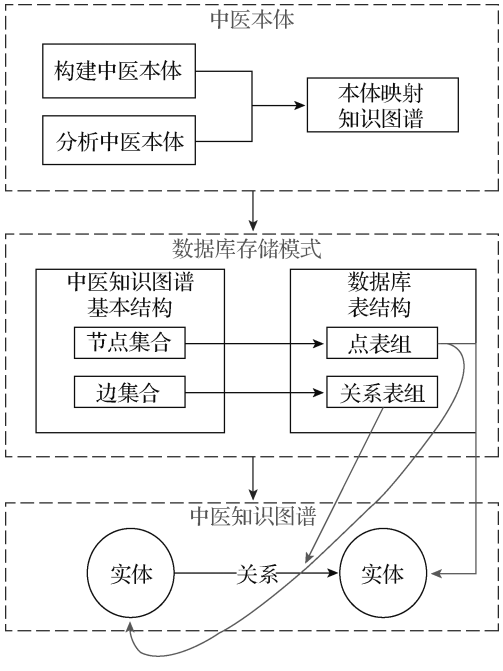

图3 知识图谱构建过程图

4.4 知识图谱的其他知识获取方式

基于本体构建的中医知识图谱如何获得更多、更全面的中医知识，需要有更多的知识源和知识获取方法进行知识的补充和知识的自我进化。目前我们已经实现的中医知识图谱知识获取方式，主要有以下几方面：

1）基于名老中医医案的知识获取

基于名老中医的经典权威医案，利用自然语言处理技术，进行中医各类词表的扩充并基于数据挖掘技术进行知识图谱中节点间关系的发现。

2）基于开源知识库的知识获取

基于网络开源知识库资源，利用爬虫技术进行中医各类词表的自动补充和完善，进行图谱的节点补充及关系的发现。

3）基于文献的知识获取

针对中医的大量文献资料，研究并实现基于模板的中医知识抽取，对知识图谱的知识进行补充，并研究知识图谱驱动下的中医知识自动抽取技术。

4）基于用户反馈的知识获取

开发若干中医知识服务应用，并基于用户的应用反馈对知识图谱的知识进行修正和补充。随着应用服务的不断加强，这一部分将是未来中医知识图谱的重要知识来源。

5 知识图谱应用探索

1）基本中医知识检索

作为中医知识图谱的基本检索服务，关于中医单个术语、多个术语、术语间关系、术语间指定关系的路径查询等基本的知识检索服务

doi:10.3772/j.issn.2095-915x.2017.01.004

都以交互式图形化的方式展示出知识图的检索结果。

2）医案分析

根据医案的信息，分析医案的临床诊断路径，帮助理解医案的辨证论治思路。

3）辅助诊疗

根据症状信息，基于知识图谱，结合多种中医辨证方法，进行辨证论治策略的推荐。

4）名老中医个性化知识分析

在名老中医医案驱动下，基于知识图谱进行渐进式名老中医学术思想的发现，为进一步的中医经验传承和临床知识总结进行探索。

6 总结

本文基于中医本体提出了中医核心知识图谱的表示和构建方法，并对中医知识图谱的知识获取和应用进行了探索。中医作为中华民族的大智慧，如何利用知识工程技术更好地表达和应用中医知识，让中医知识更好地传承和为大众服务还有很多工作要做。

参考文献

[1] 于彤，刘静，贾李蓉，等．大型中医药知识图谱构建研究 [J]. 中国数字医学，2015(3): 80-82.

[2] 阮彤，孙程琳，王昊奋，等．中医药知识图谱构建与应用 [J]. 医学信息学杂志，2016, 37(4): 8-13.

[3] 贾李蓉，刘静，于彤，等．中医药知识图谱构建 [J]. 医学信息学杂志，2015, 36(8): 51-53.

[4] 顾芳．多学科领域本体设计方法的研究 [D]. 北京：中国科学院研究生院（计算技术研究所），2004.

[5] 朱文博，李爱平，刘雪梅．基于本体的冲压工艺知识表示方法研究 [J]. 中国机械工程，2006, 17(6): 616-620.

[6] 艾丹祥．基于本体论的知识检索研究 [D]. 武汉：武汉大学，2004.

[7] 刘建炜，燕路峰．知识表示方法比较 [J]. 计算机系统应用，2011, 20(3):242-246.

[8] 王昊奋．知识图谱技术原理介绍 [EB/OL]. [2016-10-11]. http://www.36dsj.com/archives/39306.

[9] 李文哲．知识图谱的应用 [EB/OL]. [2016-11-01]. http://mp.weixin.qq.com/s?__biz=MzA3NTU3Mjg0NQ==&mid=401074459&idx=1&sn=3f20d4aae0fb97b6f231615f7db44c7a&3rd=MzA3MDU4NTYzMw==&scene=25#wechat_redirect.

[10] 果壳网．谷歌知识图谱功能带来的是什么？[EB/OL]. [2016-11-01]. http://www.guokr.com/article/203344/.

[11] 袁磊，张浩，陈静，陆剑峰．基于本体化知识模型的知识库构建模式研究 [J]. 计算机工程与应用，2006(30): 64-68,104.

[12] Protégé 官　网 [EB/OL]. [2016-11-08]. http://protege.stanford.edu/.