

项目编号

湖南省大学生创新创业训练计划项目

(创新训练项目)

申请书

项 目 名 称: 基于知识图谱的中医养生知识智能问答系统
构建研究与实现

项目实施 日期: 2020 年 7 月 1 日 — 2022 年 6 月 30 日

负 责 人: 李琳菊

联 系 方 式:

所 属 学 院: 信息科学与工程学院

学 号:

专 业 班 级: 2018 级信息管理与信息系统班

指导教师 / 邮箱: 韦昌法 /
徐宏宁 /

申 报 日 期: 2020.06.22

湖南中医药大学

填写说明

1、本申请书所列各项内容均须实事求是，认真填写，表达明确严谨，简明扼要。

2、申请人可以是个人，也可为创新团队，首页只填负责人，“项目编号”一栏不填。

3、本申请书为大 16 开本（A4），左侧装订成册。可网上下载、自行复印或加页，但格式、内容、大小均须与原件一致。

4、负责人所在学院认真审核，经初评和答辩，签署意见后，将申请书（一式两份）报送湖南中医药大学大学生创新创业训练计划领导小组办公室。

一、基本情况

项目名称	基于知识图谱的中医养生知识智能问答系统构建研究与实现						
所属学科	学科一级门：	工学		学科二级类：	计算机科学与技术		
申请金额	10000 元		起止年月	2020 年 7 月至 2022 年 6 月			
负责人姓名	李琳菊	性别	女	民族	苗族	出生年月	2001 年 11 月
学号		联系电话	宅：无 手机：				
指导教师	韦昌法,徐宏宁	联系电话	宅：手机：				
负责人曾经参与科研的情况	(1)大一上学期用 CodeBlocks 开发了一个中药管理系统，使用户能获知中药的习性、库存数量和市场价格。(2)大一下半期用 C 语言开发了校园导航系统，使新生能更快地熟悉校园。(3)大二上学期用 Java 开发了聊天系统，通过多线程支持多人聊天。(4)大二下学期用 Android Studio 开发了一个记账本，用户可进行记账和支出预算。在大二期间担任了班级团支书，具有较强的团队组织能力。						
指导教师承担科研课题情况	(1) 韦昌法副教授，主持湖南省自然科学基金面上项目 1 项、教育部科技发展中心科研创新基金课题 1 项、湖南省教育厅科学研究项目 2 项，指导学生参加程序设计竞赛和人工智能创新创业大赛并获奖。(2)徐宏宁老师，在百度、美团等知名互联网企业工作近 8 年，参与了百度搜索数据流平台项目、用户行为分析大数据项目、互联网内容分布式缓存项目和美团外卖配送项目等多个大型项目，熟悉互联网企业软件项目开发流程、开发模式及所用的一流技术。						
指导教师对本项目的支持情况	(1) 指导学生研究构建中医养生知识图谱。 (2) 指导学生研究构建基于知识图谱的中医养生知识智能问答系统。 (3)指导学生依照软件项目开发流程进行系统需求分析、设计、开发和测试。 (4)指导学生撰写科研论文和研究报告。 (5)在项目实施过程中重点培养学生的创新思维、科研能力、动手开发能力和团队协作能力。						
项目组成员主要成员	姓名	学号	专业班级		所在学院		项目中的分工
	李亚春		2018 级信息管理与信息系统班		信息科学与工程学院		知识图谱构建
	刘焱琪		2018 级医学信息工程班		信息科学与工程学院		系统构建研究
	梁玉婷		2018 级信息管理与信息系统班		信息科学与工程学院		系统开发
	惠君俊		2019 级信息管理与信息系统班		信息科学与工程学院		系统测试

二、 立项依据（可加页）

（一） 项目简介

中医养生保健，是指在中医理论指导下，通过各种方法达到增强体质、预防疾病、延年益寿目的的保健活动。中医养生保健在我国有着悠久的历史 and 深厚的文化积淀。随着我国经济的持续快速发展，人民群众的生活水平大幅度提高，对健康长寿和高质量生活的追求进一步提升，对中医养生知识服务的需求进一步旺盛。但是，目前在中医养生知识供需方面存在矛盾，一方面人民群众普遍需要了解中医养生保健知识，另一方面人民群众缺少准确、可靠、便捷的中医养生保健知识来源。介绍中医养生保健知识的书籍虽然不少、但是质量上良莠不齐，很多群众通过上网搜索、阅读微信公众号文章等方式获得中医养生保健知识，但是无法确认这些知识的可信度，这严重影响了广大群众利用中医养生知识开展自我防病、养生保健的积极性和成效。

构建基于知识图谱的中医养生知识智能问答系统，从正规渠道采集可信的中医养生保健知识，为广大群众提供可信、可靠、便捷的中医养生知识服务，对提升广大群众的养生保健能力、守护百姓健康具有重要的作用和意义。

（二） 研究目的

本项目拟利用人工智能技术从多元异构的中医养生知识来源中采集中医养生保健知识，构建中医养生知识图谱，进而构建基于知识图谱的中医养生知识智能问答系统，为用户提供可信、可靠、便捷的中医养生知识查阅服务和智能问答服务，以提升广大人民群众中医养生保健能力。

（三） 研究内容

本项目的研究内容如下：

- (1) 研究如何利用数据采集技术从目前主流的中医养生学教材、著作以及权威可靠的中医药网站和数据库中高效采集中医养生知识。
- (2) 基于上一步所采集的中医养生知识，研究梳理中医养生知识体系，构建中医养生知识数据库。
- (3) 研究构建中医养生知识图谱的结构模型，确定中医养生知识图谱中的基本类别、属性和语义关系。
- (4) 针对多元异构的中医养生知识来源，研究相应的知识抽取方案，获取命名实体和关系；针对不同数据来源中获取的命名实体和关系，研究相应的知识融合方

案。

(5) 研究中医养生知识图谱表示与存储方案，实现中医养生知识图谱的构建和存储。

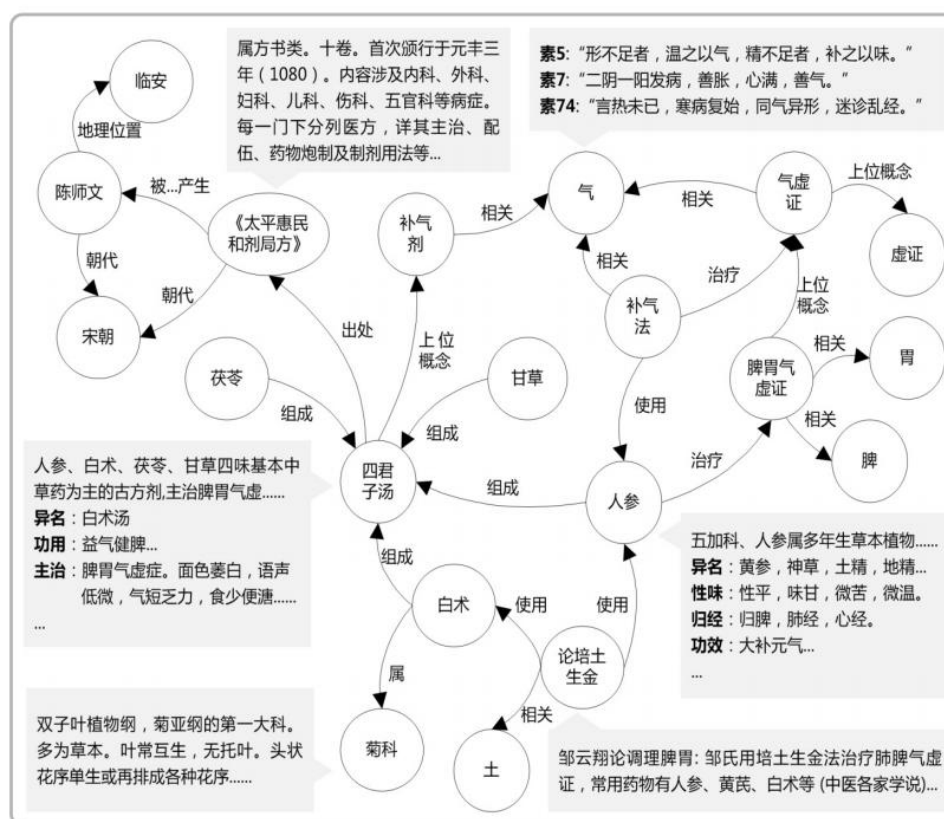
(6) 研究基于知识图谱的中医养生知识智能问答系统构建方案。

(7) 研究设计并实现基于知识图谱的中医养生知识智能问答系统，为用户提供可信、可靠、便捷的中医养生知识查阅服务和智能问答服务。

(四) 国、内外研究现状和发展动态

(1) 中医药相关知识图谱研究现状

知识图谱是人工智能领域的研究热点，它具有知识语义化、易关联、易扩充、可视化等优点，为中医药领域知识的关联、整合与利用提供了理想的技术手段。知识图谱是从语义网演化而来，被用来描述实体之间的关系，知识图谱的表达形式为三元组，即 $G = (E, R, S)$ ，其中 $E = \{e_1, e_2, \dots, e_n\}$ 表示图谱中所有实体的集合； $R = \{r_1, r_2, \dots, r_m\}$ 表示图谱中的实体之间的所有关系的集合； $S \in E \times R \times E$ 表示图谱中的三元组。中医药知识图谱能够以图谱的形式对中医药领域知识进行可视化表示，支持知识关联和扩充，为中医药教学、辅助诊疗等工作提供知识服务。下图是研究者构建的一个中医药知识图谱示例：



谷歌公司于 2012 年 5 月正式提出了知识图谱的概念。自谷歌发布知识图谱以来,多种图谱涌现出来,其中包括有 Google Graph、DBpedia、HowNet 等。近年来已有学者在中医药知识图谱构建方法与标准化流程多方面进行了尝试和探索。于彤等提出以中医药学语言系统为框架,以中医药领域现有的术语和数据库资源为内容,提出构建大型知识图谱的构想,并进行了探索和实践,但尚未实现中医药知识资源的有效整合及全面、及时、可靠的知识服务^[1]。贾李蓉等以中医知识图谱为例,从数据来源、研究内容、图形化展示等方面探讨如何构建中医知识图谱,但其应用尚局限于浏览检索方面,对多种数据资源间的映射及数据元等标准未进行详细论述^[2]。张德政等提出了基于本体的中医核心知识图谱表示及其构建方法,为中医知识图谱的构建提供了较系统的方法流程,但对多源数据的获取技术以中医医师临床实际诊疗数据的研究未进行深入研究^[3]。

(2) 智能问答系统研究现状

智能问答系统对于输入的自然语言问题能够给出简短、精确的答案,被预见为下一代互联网信息服务的基本形式^[4]。按问答系统数据的不同组织形式进行分类可将问答系统分为三类:基于半结构化知识的问答系统^[5-6]、基于文本的问答系统^[7-8]和基于问题-答案的问答系统^[9]。随着知识图谱技术的发展,以知识图谱这一新形式的半结构化数据为基础的问答系统迎来了新的机遇。

王文辉等设计并实现了基于 WordNet 计算予语义相似度算法的英文问答系统,能较为准确地回答小学英语问题,问答系统为“问题-答案”的格式^[10]。

崔万云等通过爬取 Yahoo 现在问答数据并生成大量的问句模板,并基于模板匹配结合概率模型进行问句语义理解,该方法是目前可行性和实验效果较好的方法,但主要是针对英文通用领域问答系统数据进行实现^[11]。

(3) 基于知识图谱的中医养生知识智能问答相关研究现状

目前,虽然已经有不少研究者开展了中医药领域的知识图谱相关研究和应用工作,但是基于知识图谱构建中医养生知识智能问答系统的研究尚未见详细报道。

综上所述,本项目拟利用人工智能技术从多元异构的中医养生知识来源中采集中医养生保健知识,构建中医养生知识图谱,进而构建基于知识图谱的中医养生知识智能问答系统,为用户提供可信、可靠、便捷的中医养生知识查阅服务和智能问答服务,对提升广大群众的养生保健能力、守护百姓健康具有重要的作用和意义。

参考文献:

- [1] 于彤,刘静,贾李蓉,张竹绿,杨硕,刘丽红,李敬华,于琦.大型中医药知识图谱构建研究[J].中国数字医学,2015,1003:80-82.
- [2] 贾李蓉,刘静,于彤,董燕,朱玲,高博,刘丽红.中医药知识图谱构建[J].医学信息学杂志,2015,3608:51-53+59.
- [3] 张德政,谢永红,李曼,石川.基于本体的中医知识图谱构建[J].情报工程,2017,301:35-42.
- [4] 李舟军,李水华.基于 Web 的问答系统综述[J].计算机科学,2017,4406:1-7+42.
- [5] Faezeh Eskandari,Mehdi Shafieian,Mohammad M. Aghdam,Kaveh Laksari. A knowledge map analysis of brain biomechanics: Current evidence and future directions[J]. Elsevier Ltd,2020,75.
- [6] Alizia Mantovani,Fabrizio Piana,Vincenzo Lombardo. Ontology-driven representation of knowledge for geological maps[J]. Elsevier Ltd,2020,139.
- [7] 李彦志,朱红梅.基于 Cytoscape.js 的 Neo4j 农药知识图谱 Web 可视化研究[J].软件,2020,41(04):10-13.
- [8] 胡欢,云红艳,贺英,张秀华.半自动构建扶贫领域知识图谱工具的研究[J].计算机与数字工程,2019,47(08):1961-1965+2055.
- [9] 王昊奋,漆桂林,陈华钧.《知识图谱:方法、实践与应用》[J].自动化博览,2020,3701:7.
- [10] 王文辉,吴敏华,骆力明,刘杰.基于相似度算法的英语智能问答系统设计与实现[J].计算机应用与软件,2017,3406:62-68.
- [11] 陈垚亮,洪骥,崔万云,肖仰华.BWAPlus:一个基于频繁序列的下一代基因比对工具[J].计算机研究与发展,2011,48S3:391-394.

（五）创新点与项目特色

本项目的创新点：本项目选题新颖，基于知识图谱构建技术研究构建中医养生知识图谱，基于中医养生知识图谱和自然语言处理技术研究构建并实现中医养生知识智能问答系统，为广大群众提供可信、可靠、便捷的中医养生知识服务，对提升广大群众的养生保健能力、守护百姓健康具有重要的作用和意义。

本项目的特色：作为大学生创新训练计划项目，本项目由学生团队在老师的指导下自主进行项目设计、项目实施和管理、完成项目结题，开发实现的基于知识图谱的中医养生知识智能问答系统将为用户提供可信、可靠、便捷的中医养生知识查阅服务和智能问答服务。本项目团队还计划参加大学生程序开发类竞赛，进一步推广本项目的研究成果。

（六）技术路线、拟解决的问题及预期成果

1. 项目的技术路线和拟解决的问题

(1) 查阅大量相关文献，了解国内外研究现状，咨询指导老师意见，确定研究项目：基于知识图谱的中医养生知识智能问答系统构建研究与实现。

(2) 查阅文献并咨询中医专家，研究利用数据采集技术从目前主流的中医养生学教材、著作以及权威可靠的中医药网站和数据库中高效采集中医养生知识。

(3) 基于上一步所采集的中医养生知识，研究梳理中医养生知识体系，构建中医养生知识数据库。

(4) 研究构建中医养生知识图谱的结构模型，确定中医养生知识图谱中的基本类别、属性和语义关系。

(5) 针对多元异构的中医养生知识来源，研究相应的知识抽取方案，获取命名实体和关系；针对不同数据来源中获取的命名实体和关系，研究相应的知识融合方案。

(6) 研究中医养生知识图谱表示与存储方案，实现中医养生知识图谱的构建和存储，拟采用 Neo4j 来构建和存储中医养生知识图谱。

(7) 研究基于知识图谱的中医养生知识智能问答系统构建方案，拟采用指导教师团队所熟悉的企业软件项目开发流程、开发模式和一流技术来进行系统构建设计和开发实现。

(8) 研究设计并开发实现基于知识图谱的中医养生知识智能问答系统，拟采用

自然语言处理技术来实现知识智能问答服务，为用户提供可信、可靠、便捷的中医养生知识查阅服务和智能问答服务。

(9) 最后进行项目结题工作。

2. 项目的预期成果

(1)构建中医养生知识图谱，开发实现基于知识图谱的中医养生知识智能问答系统，完成研究报告。

(2)在国家级期刊上发表研究论文至少 1 篇。

(3)培养锻炼本项目 5 名本科生的科研能力、知识图谱构建能力和程序开发能力，带动更多学生积极申报、参与大学生创新创业训练计划项目。

(七) 项目研究进度安排

第一年度(2020.07-2021.06)

(1) 2020.07-2020.08: 查阅相关文献资料，研究完善该创新训练计划项目实施方案。

(2) 2020.09-2020.12: 咨询中医专家，研究利用数据采集技术从目前主流的中医养生学教材、著作以及权威可靠的中医药网站和数据库中高效采集中医养生知识。研究梳理中医养生知识体系，构建中医养生知识数据库。

(3) 2021.01-2021.06: 研究构建中医养生知识图谱的结构模型，确定中医养生知识图谱中的基本类别、属性和语义关系，研究相应的知识抽取方案和知识融合方案。研究中医养生知识图谱表示与存储方案，实现中医养生知识图谱的构建和存储。

第二年度(2021.07-2022.06)

(1) 2021.07-2021.11: 研究基于知识图谱的中医养生知识智能问答系统构建方案，研究设计并开发实现基于知识图谱的中医养生知识智能问答系统。

(2) 2021.12-2022.03: 整理项目研究成果，撰写研究论文并投稿。

(3) 2022.04-2022.06: 撰写研究报告，准备项目结题。

（八）已有基础

1. 与本项目有关的研究积累和已取得的成绩

（1）指导教师韦昌法副教授指导学生团队开展了校级大学生研究性学习和创新性实验计划项目“中医诊断学辨证知识图谱构建研究与实践”，研究设计并开发实现了“基于知识图谱的中医诊断学辨证知识咨询平台”，指导学生撰写了科研论文“中医诊断学辨证知识图谱构建研究与实践”，已被《医学信息学杂志》录用。

论文录用通知

周海, 彭斐灵, 韦昌法 先生/女士:

经编辑部初审, 专家审稿、同行评议, 您(们)撰写的“
中医诊断学辨证知识图谱构建研究与实践”(论文编号: 2020
04070265, 作者单位: 湖南中医药大学)已被《医学信息学杂
志》录用。拟于近期发表, 版面费缴纳另行通知。

此致



指导教师韦昌法副教授还指导学生团队参加了 2019 年湖南省第 15 届大学生计算机程序设计竞赛, 并荣获二等奖。



指导教师徐宏宁老师在百度、美团等知名互联网企业工作近 8 年，参与了百度搜索数据流平台项目、用户行为分析大数据项目、互联网内容分布式缓存项目和美团外卖配送项目等多个大型项目，熟悉互联网企业软件项目开发流程、开发模式及所用的一流技术。

(2) 本项目学生团队已系统学习了 C 语言程序设计、Java 程序设计、Web 开发、移动 APP 开发等课程并开发实现了中药管理系统、校园导航系统、聊天系统、记账本 APP 等多个小型系统，积累了一定的软件开发经验，目前正在积极学习知识图谱构建技术和 Neo4j 等知识图谱构建工具，为本项目的实施奠定了扎实的基础。

2. 已具备的条件，尚缺少的条件及解决方法

本项目将依托湖南省 2011 数字中医药协同创新中心、国家中医药管理局重点学科中医药信息学等已有的研究平台来开展研究工作。本项目指导教师所归属的科研团队在中医药信息化领域具有强大的科研实力和丰富的研究经验，拥有先进的软硬件实验环境，为本项目的研究提供了扎实的基础条件。

目前，尚缺少的实验条件主要是缺乏权威可靠的中医养生知识数据来源，本项目团队将利用数据采集技术从目前主流的中医养生学教材、著作以及权威可靠的中医药网站和数据库中高效采集中医养生知识。

三、 经费预算

开支科目	预算经费 (元)	主要用途	阶段下达经费计划 (元)	
			前半阶段	后半阶段
预算经费总额	10000			
1. 业务费				
(1) 计算、分析、测试费	0			
(2) 能源动力费	0			
(3) 会议、差旅费	0			
(4) 文献检索费	1500	购买检索打印相关文献资料等	1000	500
(5) 论文出版费	7000	发表论文等	3500	3500
2. 仪器设备购置费	1500	购买网络虚拟主机和域名等	1000	500
3. 实验装置试制费	0			
4. 材料费	0			
学校批准经费				

四、 指导教师意见

该项目选题新颖、创新性强、能较好地培养和锻炼学生的创新能力和研究能力。李琳菊等同学在软件系统开发和知识图谱构建等方面掌握了一定的理论基础和较强的动手实践能力。本人同意李琳菊等同学申报“基于知识图谱的中医养生知识智能问答系统构建研究与实现”这一大学生创新训练计划项目，并将在该项目的研究过程中给予悉心指导。

导师（签章）：

年 月 日

五、 院系大学生创新创业训练计划专家组意见

专家组组长（签章）：

年 月 日

六、 学校大学生创新创业训练计划专家组意见

负责人（签章）：

年 月 日

七、 大学生创新创业训练计划领导小组审批意见

负责人（签章）：

年 月 日