基于知识图谱的中医古方交互式检索研究*

叶丁菱[1]　牟丽君[2]　许鑫[1]
（1. 华东师范大学经济与管理学院，上海 200241；2. 山东艺术学院党委办公室，济南 250300）

摘要：以中医古方为研究对象，利用命名实体识别、关系抽取等构建基于深度学习的中医古方知识图谱，并通过Neo4j图数据库实现中医古方知识与检索结果的可视化。在此基础上结合交互式检索技术，设计人机交互式检索架构，实现中医古方知识检索、知识问答、知识浏览以及知识推理。利用知识图谱对中医古方进行知识结构重组，结合交互式检索增强知识关联和提升检索效率，为中医古方相关知识组织和知识服务提供数据支持及方法借鉴。

关键词：知识图谱；命名实体识别；交互式检索；中医古方
中图分类号：G203　DOI：10.3772/j.issn.1673-2286.2024.02.003
引文格式：叶丁菱，牟丽君，许鑫. 基于知识图谱的中医古方交互式检索研究[J]. 数字图书馆论坛，2024，20（2）：24-33.

中医作为我国传统文化的瑰宝，是历代医家临床诊疗思想的结晶。中医古方作为中医学发展的重要知识成果，承载了中华民族数千年的宝贵医学知识和临床诊疗经验。深入研究和解析中医古方的辨证论治，把握古方遣药思路，对于促进古方新解新用具有重要意义。《中医药发展战略规划纲要（2016—2030年）》《中共中央 国务院关于促进中医药传承创新发展的意见》等文件的相继印发，将中医药发展上升为国家战略，传承精华、守正创新的中医药现代化建设正当其时。在大数据和人工智能背景下，应用现代信息技术，推动“智能+中医”，对中医古方进行知识挖掘和利用，成为中医药数智化发展的重要切入点之一。

中医古方的专业词汇、表达方式等具有独特性，使其在数字化背景下的知识传承与传播亟需经过规范化和结构化处理。为实现直接高效的中医知识组织与知识挖掘，中医药领域开始探索应用知识图谱对中医方剂、证候、疾病关系等用药规律、证治规律进行研究。基于此，本文以中医古方为研究对象，利用命名实体识别等技术，结合主题词表构建中医古方知识图谱，对中医古方数据进行知识组织。与此同时，引入人机交互式检索，利用用户认知结构提高中医古方知识图谱检索效率和检索质量。本文通过深层挖掘与利用知识图谱中的语义信息，推动中医古方的知识表示、检索、可视化、再发现等，为实现辅助诊疗、智能问答、智慧医疗等中医知识服务提供支持。

1　中医古方知识图谱相关研究

基于中医药概念体系繁杂、知识总量巨大、知识碎片化等现实，中医药数智化建设亟需合理高效的知识组织和管理方法。目前中医古方知识图谱研究较少，相关研究多对中医古籍等进行知识组织、图谱构建和知识检索。

收稿日期：2023-11-24
*本研究得到2021年度国家社会科学基金重大项目“文化遗产智慧数据资源建设与服务研究”（编号：21&ZD334）资助。

1.1　中医药知识图谱相关研究

国内外学者以中医药概念体系为核心，从构建方法、临床应用、流派传承等不同角度对中医药概念性知识进行组织与管理。在图谱构建方法上，张卫东等[1]以本体结合辨证论治，实现中医古籍《金匮要略》的知识图谱构建与可视化。张向先等[2]结合元数据与本体，从著录特征与内容特征角度进行敦煌吐鲁番医药文献本体构建，完成本体与知识图谱映射，实现医药文献的语义查询。在临床决策支持上，相关学者针对乙型肝炎、糖尿病、肾病等专科专病构建了专病本体、知识图谱或辅助诊疗体系[3-5]，部分学者基于中医医案、电子病历等特殊数据源构建方剂、医案类图谱并集成与推理诊疗经验[6-7]。翟兴等[8]利用知识图谱的知识融合与推理技术，从健康养生角度切入，构建了中医领域知识图谱和知识服务体系。在学术流派传承上，学者主要对古今中医名家、经典、诊疗方法之间的相关和继承关系进行梳理，分析中医古代诊疗的知识演化。凌天等[9]收集整理了1368—1912年明清浙派中医医家的地域、流派、医术特色等数据集，为深层挖掘分析浙江中医医家、医学著作、学术流派传承，以及搭建中医医家知识库、可视化知识图谱等提供数据支撑。中医药知识图谱以语义网络为核心，建立概念之间的语义关系，囊括相关词、定义、属性值、注释等丰富的信息资源，实现对中医药领域知识的利用与再发现。

1.2　中医术语主题词表相关研究

医学术语是医学领域利用特定文字来表述或限定专业概念的符号，利于规范、分类和挖掘领域的核心知识。2019年5月世界卫生大会通过《国际疾病分类第十一次修订本（ICD-11）》，ICD-11首次纳入以中医药为代表的传统医学章节并于2022年1月生效[10]。美国国立医学图书馆构建的一体化医学语言系统（Unified Medical Language System，UMLS）收录了300多万个生物医学概念，被广泛应用于不同健康医疗信息系统、文献资源与计算机系统的语义互操作和知识内容关联[11]。由中国中医科学院中医药信息研究所牵头，全国13家科研院所参与编制的中医药学语言系统（Traditional Chinese Medicine Language System，TCMLS）[12]是参照UMLS，考虑中医领域学科特点和概念使用习惯，利用本体的核心思想编研而成的我国大型中医药领域术语体系[13]，其定义了96种实体类型、58种实体关系和127种语义类型，以概念为核心对中医术语进行系统梳理和精准表达[14]，建立网络化的信息组织框架，为中医药领域知识的组织、利用提供了体系支撑。

1.3　交互式检索相关研究

交互式检索在中医药领域应用尚不广泛，但其能在检索结果不佳的情况下，基于用户信息反馈不断修正检索结果，直至接近用户检索意图[15]。对于专业性较强的中医药领域，交互式检索基于增强人机信息交互、关注用户反馈信息、修正和拓展用户查询结果等特性，能有效提高知识图谱的知识检索和知识服务的效率与质量。国内外对交互式检索的研究主要聚焦于用户信息行为和认知发展、支撑技术与方法、模型构建与系统开发。Jackson等[16]使用临床记录交互式搜索数据资源，从临床文本中提取出严重精神疾病的关键症状，促进心理健康数据的再利用。Zheng等[17]探索了基于知识图谱的交互式自然语言查询技术。Zhang等[18]构建了基于视频的跨模态交互式检索网络，用于探索视频和查询内容的潜在关系。刘萍等[19]将认知建构理论运用于信息检索领域，构建了交互式信息检索模型，促进了用户认知发展。吴丹等[20]、陈乐等[21]结合眼动追踪技术构建了交互式信息检索模型，深层挖掘信息检索背后的用户信息行为和信息需求。

综上所述，中医药领域专业术语和行文体系独特，同物异名和同名异物等现象普遍存在，以知识图谱形式对其进行知识组织十分必要。然而，中医古方知识图谱构建相对欠缺，用户需求的专业性与多元性使中医古方的有效检索和利用尚存阻碍。因此，本文试图构建中医古方知识图谱，通过命名实体识别和实体对齐等方式组织中医古方知识数据，在此基础上加入控制词表，引入交互式检索机制，根据用户交互信息从需求侧提高用户检索效率，并在用户认知“同化—顺应”过程中补充中医药领域有关主题词表，以期助力语义级知识问答和辅助诊疗开方，实现对中医古方数据的知识挖掘和再发现。

2 中医古方知识图谱构建

中医古方知识图谱所包含的概念与范围可控，研究采用自顶向下的方法进行中医古方知识图谱构建，利用BERT-BiLSTM-CRF模型进行命名实体识别，复用TCMLS框架中的语义关系定义实体关系，并适当补充完善，并将抽取的实体及关系以三元组形式导入Neo4j图数据库，存储、检索并可视化中医古方数据特征，构建中医古方知识图谱。

2.1 数据获取

数据来自博览医书数据库（https://www.imedbooks.com/）[22]，该数据库由中国中医科学院权威数据支持，全面收集目前已经公开的古方文献资料，对方剂药物的组成、功效、主治、用法用量有详细的解释说明。近年病毒感染肆虐，中医药发挥了重要的应急救助作用。研究选取包含有肺炎、发热、畏寒、咳嗽、身疼、呕吐等关键词的相关方剂，通过Python程序爬取5 000条实验数据，其中包含国家中医药管理局发布的选自《伤寒论》《金匮要略》《千金翼方》《备急千金要方》等中医领域代表性古籍的经典名方。所获取的古方数据为半结构化数据，每个古方按药物组成、功效、主治、用法用量等字段进行介绍。因主要探究方剂组成、疾病、证候、症状之间的关系，仅爬取方剂名称、中药组成、功能主治3个字段。

2.2 本体结构

《健康信息学 中医药学语言系统语义网络框架》（GB/T 38324—2019）通过描述中医药学语言系统概念间的关系来定义语义网络的概念结构，适用于中医药学知识体系的构建[23]。该框架以概念为核心对中医术语进行体系梳理和精准表达，对中医药古方知识进行了系统而详细的划分，是中医药领域的规范化成熟本体。研究在本体设计上直接复用该语义网络框架中的部分内容，并在此基础上进行一定补充和完善。具体而言，类主要包括方剂、中药、证候、症状、疾病，类间关系包括方剂与中药的组成关系、方剂与证候的治疗关系、方剂与疾病的治疗关系、证候与症状的对应关系等，本体类型和属性如表1所示。

表1 本体类型和属性

类	属性/关系	描 述	类 别
方剂（Formula）	formula_name	方剂的规范名称	数据属性
	contain	方剂的中药组成	对象属性
	treat	方剂治疗的证候	对象属性
	cure	方剂治疗的疾病	对象属性
中药（Medicine）	medicine_name	中药的规范名称	数据属性
	other_name	中药的别名	数据属性
证候（Syndrome）	syndrome_name	证候的规范名称	数据属性
	other_name	证候的别名	数据属性
	manifest	证候导致的具体症状	对象属性
症状（Symptom）	symptom_name	症状名称	数据属性
疾病（Disease）	disease_name	疾病名称	数据属性
	other_name	疾病的别名	数据属性

2.3 知识抽取

2.3.1 数据预处理

中医古方数据为半结构化数据，其中方剂名称和中药组成经过处理后可直接导入应用，而功能主治部分的证候、症状、疾病混杂，为提高命名实体识别准确性，需对相关数据进行预处理。通过数据格式转化提取出功能主治字段，清洗其中不合法字符，并人工标注训练数据。采用命名实体识别常用的标注体系BIO方式，邀请医学专业人员对主题词表中的相关内容和400条数据进行人工标注，对有争议的标注类型通过查询专业术语词表进行规范。文本中的每个字符被标注为“B-X”“I-X”或“O”（B表示本实体的第一个字符，I表示实体的其余字符，X表示实体类型[24]，O表示其他非实体），包括证候、症状、疾病等实体类型。其中：证候表征患者在患病周期内某一阶段的病理概括，反映出症状表征下的病理变化本质；症状指患者感受到或客观存在的身体异常或某些部位的病态改变；疾病表示一定病因作用引发的机体内外环境失调，反映某疾病全过程特征和规律；非实体则指代实体之外的其他内容。以“伤寒、温病、暑病余热未清，气津两伤证。身热多汗，心胸烦闷……”为例，标注情况如表2所示。

2.3.2 命名实体识别

借鉴翟羽佳等[25]的研究，使用BERT-BiLSTM-

表2　BIO部分标注结果

文本数据	BIO标注	文本数据	BIO标注	文本数据	BIO标注
伤	B-Disease	未	O	热	I-Symptom
寒	I-Disease	清	O	多	B-Symptom
温	B-Disease	气	B-Syndrome	汗	I-Symptom
病	I-Disease	津	I-Syndrome	心	O
暑	B-Disease	两	I-Syndrome	胸	O
病	I-Disease	伤	I-Syndrome	烦	B-Symptom
余	B-Symptom	证	O	闷	I-Symptom
热	I-Symptom	身	B-Symptom		

CRF模型进行命名实体识别。有学者认为相较于其他模型，BERT-BiLSTM-CRF模型在中医古文的实体识别中更具准确性，但命名实体识别有赖于大规模的高质标注语料[26-27]。因此，鉴于中医古方专业术语复杂、语料数据基数较小，不易通过表象特征对古方思维与词汇特征进行机器学习，复杂句型句式的匹配和抽取更难等特点，为保障实验结果的可靠性，研究中全程进行人工审核和辅助。首先将半结构化中医古方数据序列输入BERT预训练语言模型，输出相应的字符特征向量，提取其中蕴含的中医药语义特征。其次，将特征输入BiLSTM模块提取上下文特征，得到完整的隐状态序列。最后，CRF模块通过转移分数对隐状态序列进行约束修正，获取标注序列，并对序列中的各个实体进行提取分类，从而完成命名实体识别。将200条古方数据以及《中医临床诊疗术语》等主题词表中的疾病、证候部分作为训练集（测试集数据100条，验证集数据100条），把4 600条未标记数据输入模型，得到标记结果。基于BERT-BiLSTM-CRF模型在中医古方数据集下得到72%的F1值，表明训练模型效果合格，模型具备一定可用性。

所得核心实体类型包括方剂名称、中药组成、证候、症状、疾病五大类，如表3所示。通过命名实体识别获取证候、症状、疾病3类实体，方剂和中药类实体则通过Excel软件的数据分列、WrapRows函数等处理获得。

2.3.3　实体关系抽取

实体关系的确定离不开词汇语义学，在知识图谱构建中不同类型实体的连接依赖实体间复杂的语义关系。基于词典的方法是实体关系抽取的重要方式，借助行业权威词典可以最大限度保证数据质量。所得中医古方所包含的实体类型只有五大类，实体间关系较为清楚明确，因此结合中医古方数据集特点选用基于词典的语义关系定义方法。

经过抽取的实体可能存在表达多样、冗余的问题，实体中有大量重复和噪声，需要开展实体对齐的工作，即将名称不同但实际上是同一实体的数据进行融合，以提高知识图谱的数据质量。在实体对齐中，主要将识别出的实体名称与《中医古籍后控词表》《中医临床诊疗术语》《中医病证分类与代码》《中医临床常见症状术语规范》等词表中的规范名进行匹配，以主题词表的规范名称为标准，将主题词表中的别名和其他相似度较高的实体名称作为实体的别名，如将“肺实热证”“肺火证”“邪热壅肺证”统一为“肺热炽盛证”。

通过对齐的实体数据以及人工处理未覆盖的实体、去除明显错误的实体，完成知识的融合，最终整理出13 311条实体数据。将对齐前后的实体进行匹配，为每个实体赋予唯一标识，各类实体通过定义的实体间关系形成“实体-关系-实体”三元组，将最终数据导入Neo4j数据库构成知识图谱。实体类型及实体关系数量如表4所示，中医古方实体关系详情如表5所示。

表3　命名实体识别示例

示 例	证 候	症 状	疾 病	方剂名称	中药组成
瘀热蓄结下焦之症。太阳病不解，热结膀胱，少腹胀满，大便黑，小便利，燥渴，其人如狂，至夜发热	瘀热蓄结下焦	热结膀胱、少腹胀满、大便黑、小便利、燥渴、如狂、发热	太阳病	桃核承气汤	桃仁、大黄、桂枝、甘草、芒硝

表4　对齐后的实体类型及实体关系数量

实体类型	实体数量/个	关系类型	关系数量/个
方剂	4 600	方剂-中药	36 419
中药	3 890	方剂-证候	903
疾病	2 027	方剂-疾病	2 991
证候	481	证候-症状	2 785
症状	2 193	疾病-别名	106
别名	120	证候-别名	14

2.4　知识存储与可视化

在对中医古方数据进行信息抽取和知识融合的基础上，将实体和关系导入Neo4j图数据库，采用Neo4j自带的load csv方式，将数据按行存储成csv文件，通过Cypher语言调用csv文件，构成知识图谱，可视化地呈现古方数据中包含的各类实体及实体之间的关系。

表5　中医古方实体关系示例

实体1	实体1类型	关 系	实体2类型	实体2
桃核承气汤	方剂	contain	中药	桃仁
甘草泻心汤	方剂	contain	中药	黄芩
竹叶石膏汤	方剂	treat	证候	气津两伤证
猪苓汤	方剂	treat	证候	水热互结证
麻黄汤	方剂	cure	疾病	风寒感冒
风寒表实证	证候	manifest	症状	发热
阳虚水泛证	证候	manifest	症状	畏寒肢厥
中风	疾病	other_name	别名	偏枯、卒中

以麻黄汤为例，麻黄汤及其对应的中药、证候、症状等如图1所示。知识图谱包含了方剂（麻黄汤）、中药（甘草、麻黄、杏仁等）、证候（外感风寒表实证）、症状（恶寒发热、舌苔薄白等）4种实体类型，以及contain、treat、manifest 3种关系，各实体节点通过相关关系连接。由此间关系可知，应用中药方剂治疗外感风寒表实证，可以以麻黄、杏仁、甘草入药。

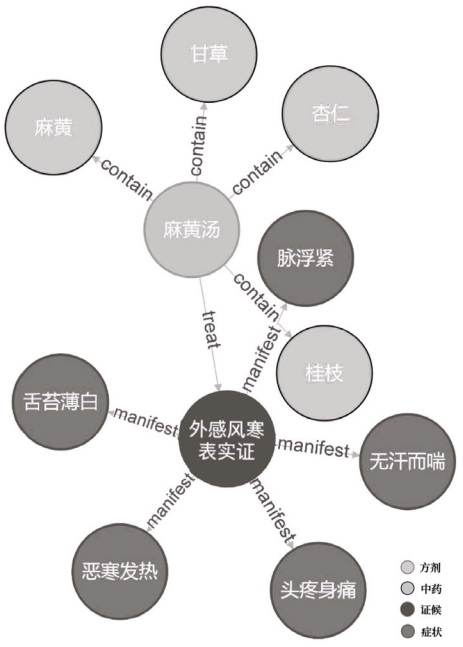

图1　麻黄汤部分知识图谱

3　基于中医古方知识图谱的交互式检索与应用

交互式信息检索更加注重用户与系统的交互过程，通过改变用户的认知结构，提高检索的查全率和查准率。为增强信息检索系统交互功能，构建基于中医古方知识图谱的人机交互式检索机制，综合查询扩展（Query Expansion，QE）和相关反馈技术的核心思想，使得用户在输入检索式后继续参与后续检索，尝试通过扩大检索范围、提高检索精度满足用户需求、辅助临床诊疗。

3.1　交互式检索方法

3.1.1　查询扩展

查询扩展是指用户通过浏览检索结果，积极干预查询式的重构与完善，将范围更准确的检索词和短语加入初始查询，通过相似度计算、自然语言处理等方法，将与初始查询关键词存在同义、相关关系的概念添加到初始查询当中，从而补充和扩展首次查询结果[28-29]。中

医药历经几千年发展，由于时间、空间变迁等，中医药词汇的内涵和外延发生改变，由此产生对同一事物的不同表述，但此类表述之间存在相同或相关的语义关系。用户在检索时，可能会使用与检索目标同义的入口词（非标准词）表达查询需求。如果孤立地对某个关键词进行检索，而忽视其在中医药领域中的同义关系、上下位关系、相关关系，则很有可能漏检重要概念信息。因此，在进行知识查询的时候，应当将词义扩展纳入考量，获取关键词的同义词、相关词等作为扩展词集合并加入后续检索环节，完善语义检索手段。将疾病与别名之间的对应关系加入检索式可以有效扩大检索范围，获取更多治疗方剂信息。例如，“中风”这一疾病类型在中医古方中有“卒中”“偏枯”等不同表述，在检索时若只将“偏枯”作为检索词，检索结果较少，但若在检索过程中加入疾病别名，以MATCH（p1：formula）-[r1：cure]->（p3：disease）-[r2：other_name]->（p6：alias {alias：‘偏枯’}）return p1，r1，p3，r2，p6为查询语句，则可以获得更多的方剂信息。

3.1.2　相关反馈

用户信息需求复杂多变，常存在多主题检索需求，加之普通用户难以通过输入检索式来准确表达信息需求，检索领域出现了基于查询扩展的相关反馈[30]。不同于查询扩展需要对关键词进行同义词扩展，构造新的查询式，在相关反馈中用户只需要对比检索结果与查询目的，指出其中哪些知识节点相关或者不相关，经过多次迭代后系统就能够更加准确把握查询式变化，根据用户反馈主动对结果进行调整，进而满足用户信息需求[31]。相较于查询扩展在检索前期通过调整查询式来扩展检索结果范围，相关反馈更加侧重于在检索后期通过选择检索结果来提高检索精度。

3.2　交互式检索架构设计

综合查询扩展和相关反馈技术，在收到用户相关反馈的情况下，在后续的每一次迭代中把相关反馈和查询扩展两个部分封装在一起，并在检索迭代过程中对知识图谱所用的控制词表进行修正。设计的基于中医古方知识图谱的人机交互式检索架构主要包括5个部分，如图2所示。

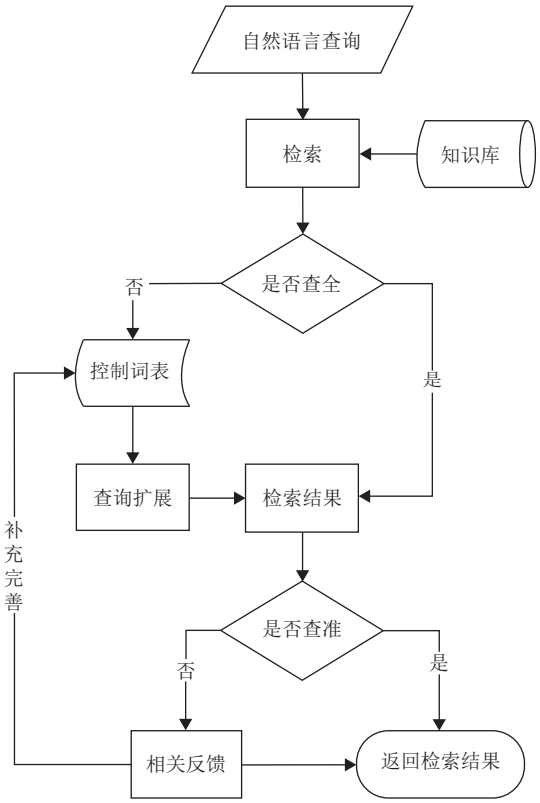

图2　基于中医古方知识图谱的人机交互式检索架构设计

（1）知识存储部分。构建的知识图谱作为系统知识库囊括了海量的中医古方知识，是实现信息存储和相关知识点展示的关键，也是检索系统的基础。用户通过该部分查看方剂知识，包括方剂名称、中药组成、疾病、证候、症状等相关信息，检索系统对用户输入的检索词通过该部分各类型实体进行比对匹配。

（2）检索部分。检索部分是用户交互的入口，用户在中医古方知识图谱检索系统中输入想要查询的方剂知识，前端将识别出的关键词信息发送给后台，后台对知识库中的节点内容与查询式中的关键词进行相似度计算，并根据计算结果将相似度较高的检索结果返回给用户，进行降序排列呈现。返回的初次查询结果是后续交互式检索的依据和基础。在交互式检索阶段，系统会根据用户的显式反馈信息，利用相关算法和检索模型对相关文档进行动态调整，在二次查询及之后迭代的每次查询中，检索结果都会结合用户反馈进行倒排，直到满足用户的信息需求。

（3）查询扩展部分。查询扩展部分可扩大检索范围，实现“查全”。针对中医药领域存在的术语体系复杂、表述模糊、难以精准检索等问题，根据《中医临床

诊疗术语》等词表将疾病、证候等实体的别名一并导入中医古方知识图谱作为受控词表，对古方数据中的非标准词汇进行规范。查询扩展的实现分为以下步骤：①用户根据初次查询结果对检索结果的进行判断，在不满意的情况下调整检索式。②以用户输入的查询式为准，对查询语句进行句法分析，提取其中的关键词，对识别出的关键词进行语义分析和词义消歧。由于研究直接通过受控词表对各实体进行异名关联，系统可以根据语义通过调用相应接口从控制词表中筛选出与关键词相关的内容，并返回扩展词集合，由用户选用相关性较高的扩展词加入二次查询。③查询扩展结果呈现。系统根据二次查询式中的关键词，利用知识库进行二次匹配，后台通过判别用户输入的实体类型确定检索的知识体系，对检索内容进行扩展，最终以知识图谱的形式返回查询扩展结果，进行可视化呈现。如果检索系统无法识别用户提交的查询式，或在已有的知识图谱库中无法匹配到相关信息，那么可以利用相关反馈进行查询扩展[32]，生成扩展词，用户根据扩展词适当调整查询式，对检索词进行适当替换，并在控制词表中补充完善。

（4）用户反馈部分。用户反馈部分可以在海量检索结果中通过相关性筛选找到满足用户信息需求的知识点，提高检索精度，实现“查准”。基于中医古方知识图谱的人机交互式检索，可以采用显式相关反馈的思想。用户反馈的实现包括以下步骤：①根据用户反馈判断结果能否满足用户的需求。如果返回结果中的前N项与信息需求高度相关，那么用户无需对检索结果进行相关性反馈；如果用户无法在海量返回结果中快速获得与需求相关的结果，那么对返回结果进行相关性判断十分必要。②用户对返回结果进行相关性判断，将每一条结果标注为相关/不相关。③系统调整与呈现检索结果。根据用户反馈信息，系统通过匹配用户需求与图谱信息，筛选出相关性较高的文档集合，对此部分信息应用相关算法，在知识库中匹配相关信息，改善查询结果。

（5）历史记录部分。历史记录存储了用户以往的检索数据，包括输入查询式、扩展词集合以及用户反馈等，用以记录和调用用户的检索历史和反馈历史。一方面，在用户无法进行显式反馈时，系统可以综合用户当前浏览时长、以往检索历史和交互行为进行检索分析；另一方面，当不同用户对同一检索内容进行查询时，系统可以调用历史记录中其他用户的反馈信息，预估当前用户可能感兴趣的内容，调整相关文档权重，返回更符合用户需求的内容。

3.3 以发热症状为例的交互式检索知识图谱应用

以“发热”为检索入口词在知识图谱中进行查询，结果如图3所示。

从图3可以发现，初次检索后系统会返回较多检索结果，其中不乏与检索需求无关的结果，这严重增加了用户的认知负荷。引入人机交互式检索后，用户出于检索目的会主动调整对检索主题的认知结构，深层加工当前信息。在内部认知发展的基础上，用户不断调整自身认知结构，形成相应的检索策略，并根据认知发展调整外部信息行为，从而迭代优化相应的检索行为。

一方面，基于查询扩展，用户可以调整检索词，重构查询式。例如在“发热”症状下，患者还有畏寒的症状，用户在认知结构完善后，调整检索关键词为“恶寒发热”。如图4所示，系统返回以“恶寒发热”为症状的相关证候和方剂，有效降低了用户认知负荷。

另一方面，基于相关反馈，用户可以对检索结果进行相关/不相关反馈。例如患者的“发热”症状明显属于“伤寒”这一证候，用户只需对“伤寒”做出相关反馈。如图5所示，系统返回以“伤寒”为证候、“发热”为症状的治疗方剂。

基于知识图谱的交互式检索相较于一般查询加深了用户对检索主题的认知和理解，合理将资源加工与用户认知相匹配，在一定程度上降低了认知负荷，有助于改善用户检索效果。

3.4 基于知识图谱的中医古方交互式检索相关应用

（1）精确的知识检索。构建的中医古方知识图谱除了以图的形式直观展示古方知识外，还可以利用Cypher语言进行目标节点和关系的查询。利用知识图谱可以实现同病异治、异病同治，使得辨证诊疗方法更加清晰，实体之间的关联更加明确，甚至能挖掘出某些不明确的实体间关系。同样为风寒感冒，有外感风寒表实证，以恶寒发热、头疼身痛、无汗而喘、舌苔薄白、脉浮紧为主要症状，可以使用麻黄汤加减进行治疗；也有

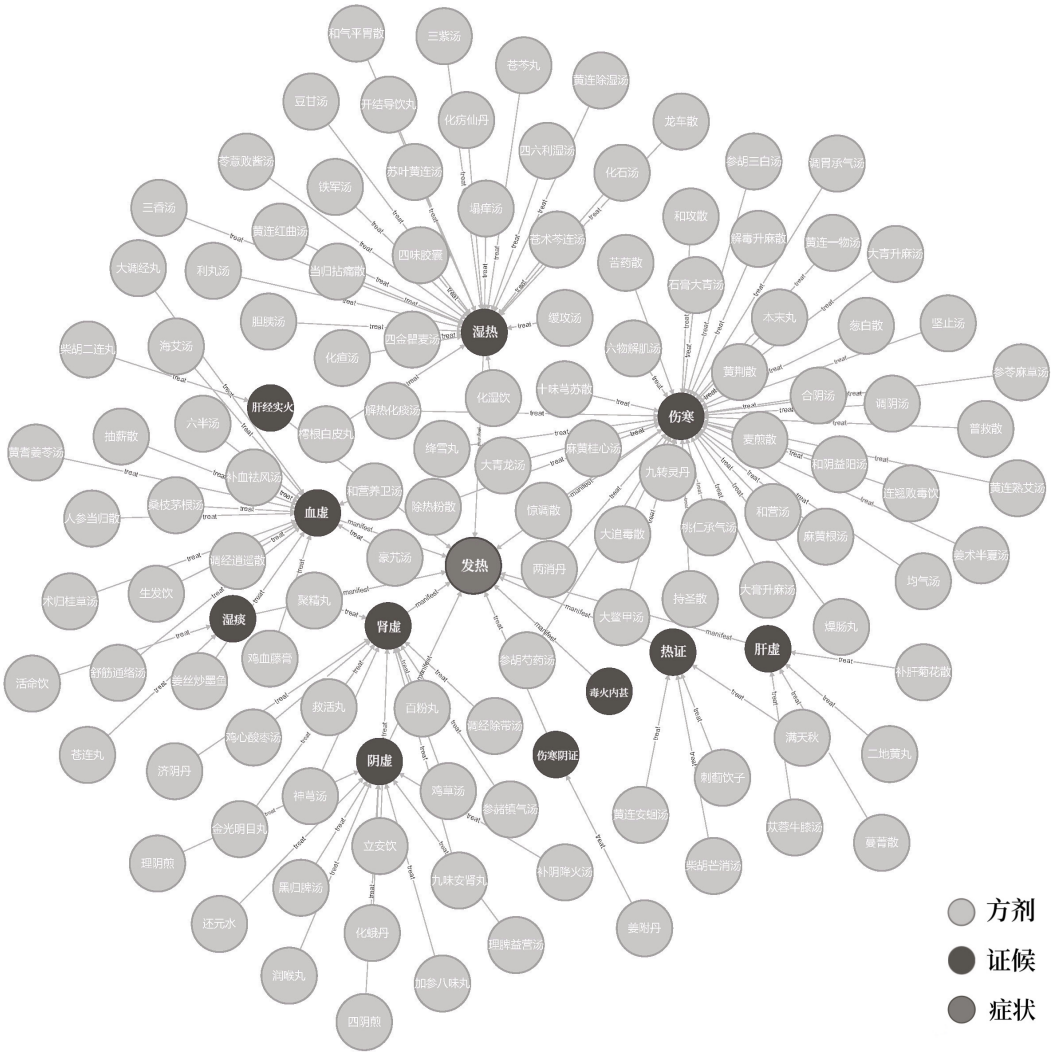

图3　“发热”症状相关知识图谱

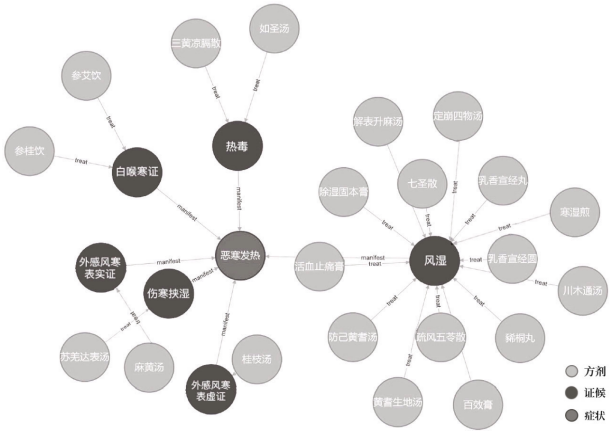

图4　“恶寒发热”有关知识图谱

外感风寒表虚证，以恶寒发热、头身疼痛、下利便溏、口不渴、舌淡苔白滑、脉浮虚、腹痛为主要症状，可以使用桂枝汤加减进行治疗。

（2）知识推理与辅助诊疗。中医古方知识图谱可以成为中医临床辅助诊疗的后台知识，其中包含的症状、证候、疾病关联关系与知识推理相结合，可以帮助医生了解病患基本症状，辨证治疗，辅助医疗诊断和开方。例如病人出现胁痛低热、心急烦躁、尿黄便秘、舌红苔黄等症状，根据知识图谱可以判断为肝郁气滞证，治疗方剂有柴胡疏肝散、四物逍遥汤、理气化瘀消肿汤、舒肝活血通经汤、舒肝破瘀通脉汤等，柴胡疏肝散的中药组成包括柴胡、香附、川芎、陈皮、枳壳、甘草。但即使是同一疾病甚至同一证候，医生也需要针对患者表现症状作出不同的诊断，对于不同的病人要随症加减。比如，如果确定病人是肝郁气滞证，且有口苦口干的症状，在开方过程中，在柴胡疏肝散中需要去除川芎，增加牡丹皮、栀子等中药。

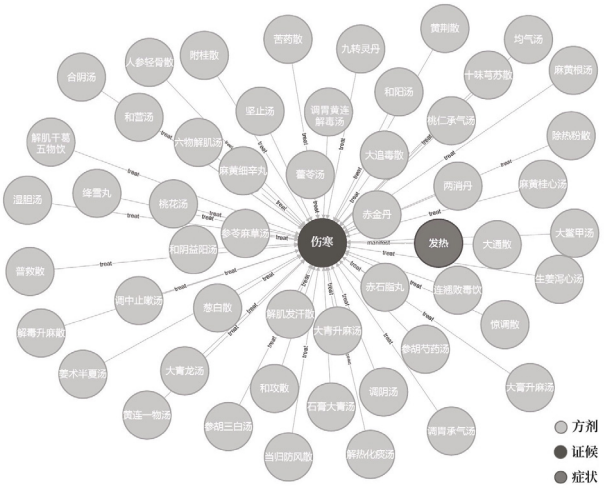

图5 “伤寒”有关知识图谱

4 结语

本文以中医古方为研究对象，一方面构建了中医古方知识图谱，结合主题词表，可视化展示中医古方中凝练的中医学理论与实践经验，满足用户获取中医知识的需求，有利于用户把握知识点之间的关联，甚至实现知识再发现，促进中医药知识的传承创新发展。另一方面将知识图谱与交互式检索相结合，实现用户与检索系统的有效交互，降低用户认知负荷，提高检索效率和检索质量，便于用户更为便捷、具象地了解和认知中医药。

本文尚存在部分局限性，可在未来研究中加以改进。一是中医药所属医疗行业严谨性较高，研究的知识抽取准确率不够高，仍需辅以人工审核，后续可以通过扩大训练集规模、融入中医领域专家观点等继续提高模型准确率。二是针对交互式检索机制，目前缺少反馈机制来提高系统表达检索者信息需求的能力，未来的研究可侧重于反馈机制的完善与应用。

参考文献

[1] 张卫东，张晓晓. 中医古籍数字资源知识组织与可视化研究：以《金匮要略》为例[J]. 情报科学，2022，40（8）：107-117.

[2] 张向先，李世钰，沈旺，等. 数字人文视角下敦煌吐鲁番医药文献知识组织研究[J]. 图书情报工作，2022，66（22）：28-43.

[3] YIN Y T，ZHANG L，WANG Y G，et al. Question answering system based on knowledge graph in traditional Chinese medicine diagnosis and treatment of viral hepatitis B[J]. BioMed Research International，2022，2022：7139904.

[4] 张玉洁，白如江，许海云，等. 融合多自然语言处理任务的中医辅助诊疗方案研究：以糖尿病为例[J]. 数据分析与知识发现，2022，6（1）：122-133.

[5] XIE J D，HE J Y，XIA P，et al. Real-world big data processing and analysis for traditional Chinese medicine 2022[J]. Evidence-Based Complementary and Alternative Medicine，2023，2023：3169031.

[6] 王成文，熊励. 基于知识图谱的突发公共卫生事件辅助诊疗研究[J]. 情报科学，2023，41（4）：164-174.

[7] 马捷，王珏，孙恒宇，等. 基于医案元数据的中医诊疗数据集构建方法与实证研究[J]. 图书情报工作，2021，65（2）：27-36.

[8] 翟兴，王涛，韩芳芳. 基于知识图谱的健康养生智能知识服务系统架构设计[J]. 信息资源管理学报，2020，10（3）：36-48.

[9] 凌天，焦阳，李露芳，等. 明清浙派中医医家数据集（1368—1912年）[J]. 中国科学数据（中英文网络版），2022，7（3）：343-350.

[10] 严世芸，胡鸿毅，黄奕然. 国际化视野下的中医药现代知识体系构建与学科建设再认识[J]. 中国大学教学，2020（4）：17-23.

[11] 李晓瑛，李军莲，李丹亚. 一体化医学语言系统及其在知识发现中的应用研究[J]. 数字图书馆论坛，2019，4（9）：24-29.

[12] CUI M，JIA L R，YU T，et al. Current status of traditional Chinese medicine language system[C]//Frontier and Future Development of Information Technology in Medicine and Education，2014：2287-2292.

[13] LIU L J，LIU L，FU X D，et al. A cloud-based framework for large-scale traditional Chinese medical record retrieval[J]. Journal of Biomedical Informatics，2018，77：21-33.

[14] ZHANG T T，HUANG Z H，WANG Y Q，et al. Information extraction from the text data on traditional Chinese medicine：a review on tasks，challenges，and methods from 2010 to 2021[J]. Evidence-Based Complementary and Alternative Medicine：ECAM，2022，2022：1679589.

[15] 刘萍，李斐雯，杨宇. 国外交互式信息检索研究进展[J]. 情报理论与实践，2017，40（5）：132-138.

[16] JACKSON R G，PATEL R，JAYATILLEKE N，et al. Natural language processing to extract symptoms of severe mental illness from clinical text：the Clinical Record Interactive Search Comprehensive Data Extraction（CRIS-CODE）project[J]. BMJ Open，2017，7（1）：e012012.

[17] ZHENG W G，CHENG H，YU J X，et al. Interactive natural

language question answering over knowledge graphs[J]. Information Sciences，2019，481（C）：141-159.

[18] ZHANG Z，LIN Z J，ZHAO Z，et al. Cross-modal interaction networks for query-based moment retrieval in videos[C]// Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval，2019：655-664.

[19] 刘萍，叶方倩，杨志伟. 认知建构视角下交互式信息检索模型研究[J]. 图书情报知识，2020（2）：93-101，122.

[20] 吴丹，刘春香. 交互式信息检索研究中的眼动追踪分析[J]. 中国图书馆学报，2019，45（2）：109-128.

[21] 陈乐，刘迎春. 基于用户需求挖掘的交互式信息检索算法设计[J]. 计算机仿真，2022，39（5）：418-422.

[22] 博览医书数据库[EB/OL]. [2023-07-19]. https://www.imedbooks.com.

[23] 中华人民共和国国家市场监督管理总局，中国国家标准化管理委员会. 健康信息学 中医药学语言系统语义网络框架：GB/T 38324—2019[S]. 北京：中国标准出版社，2019.

[24] 张云秋，汪洋，李博诚. 基于RoBERTa-wwm动态融合模型的中文电子病历命名实体识别[J]. 数据分析与知识发现，2022，6（Z1）：242-250.

[25] 翟羽佳，田静文，赵玥. 基于BERT-BiLSTM-CRF模型的算法术语抽取与创新演化路径构建研究[J]. 情报科学，2022，40（4）：71-78.

[26] 李贺，祝琳琳，刘嘉宇，等. 基于本体的简帛医药知识组织研究[J]. 图书情报工作，2022，66（22）：16-27.

[27] QU Q Q，KAN H X，WU Y T，et al. Named entity recognition of TCM text based on Bert model[C]//2020 7th International Forum on Electrical Engineering and Automation（IFEEA），2020：652-655.

[28] AZAD H K，DEEPAK A. Query expansion techniques for information retrieval：a survey[J]. Information Processing and Management，2019，56（5）：1698-1735.

[29] 黄名选，蒋曹清，卢守东. 基于词嵌入与扩展词交集的查询扩展[J]. 数据分析与知识发现，2021，5（6）：115-125.

[30] 余传明，蔡林，胡莎莎，等. 基于深度学习的查询扩展研究[J]. 情报学报，2019，38（10）：1066-1077.

[31] 徐彤阳，邓颖慧. 学术期刊APP应用中交互式检索的情景设计与技术实现[J]. 数字图书馆论坛，2019（6）：39-45.

[32] WANG J M，PAN M，HE T T，et al. A pseudo-relevance feedback framework combining relevance matching and semantic matching for information retrieval[J]. Information Processing & Management，2020，57（6）：102342.

作者简介

叶丁菱，女，博士研究生，研究方向：数字人文、科技情报、开放数据。
牟丽君，女，硕士，助理馆员，研究方向：数字人文。
许鑫，男，博士，教授，通信作者，研究方向：信息分析、科技情报、数字人文，E-mail：xxu@infor.ecnu.edu.cn。

Interactive Retrieval of Traditional Chinese Medicine Prescriptions Based on Knowledge Graphs

YE DingLing[1]　MOU LiJun[2]　XU Xin[1]
(1. School of Economics and Management, East China Normal University, Shanghai 200241, P. R. China; 2. General Committee Office, Shandong University of Arts, Jinan 250300, P. R. China)

Abstract: In this paper, we took traditional Chinese medicine prescriptions as the research object. The knowledge graph of traditional Chinese medicine prescriptions based on deep learning was constructed by using named entity recognition and relationship extraction, and the knowledge of traditional Chinese medicine prescriptions and the retrieval results were visualized by Neo4j graph database. On this basis, the human-computer interactive retrieval architecture was designed by combining the interactive retrieval technology to realize the knowledge retrieval, knowledge quiz, knowledge browsing, and knowledge reasoning of traditional Chinese medicine prescriptions. This paper used knowledge graph to reorganize the knowledge structure of traditional Chinese medicine prescriptions, and enhanced the knowledge association and retrieval efficiency by combining interactive retrieval, so as to provide data support and method reference for knowledge organization and knowledge service related to traditional Chinese medicine prescriptions.

Keywords: Knowledge Graph; Named Entity Recognition; Interactive Retrieval; Traditional Chinese Medicine Prescription

（责任编辑：王玮）