

基于知识图谱的肠造口相关并发症智能问答系统的构建研究

陈雅楠¹,陈紫婷¹,谢弋戈¹,尚浩彬²,王志稳¹,张旭¹

(1.北京大学 护理学院,北京 100191;2.北京大学 信息科学技术学院)

【摘要】 目的 构建基于肠造口并发症知识图谱的智能问答系统。**方法** (1)知识图谱构建:通过整合肠造口相关并发症指南等多种知识来源,采用本体构建七步法设计了肠造口相关并发症知识图谱的本体,之后采用自动机(aho-corasick automaton,AC)多模式匹配算法,建立造口并发症实体-关系-实体三元组,构建造口并发症知识图谱;(2)智能问答系统构建:建立肠造口相关并发症自然语言词库和特征词库,以模板匹配的方式对问题进行分类,然后形成适用于图数据库查询语言(Cypher)语法的结构化查询语言(structured query language,SQL)匹配语句,完成知识图谱查询任务,最终建立肠造口相关并发症智能问答系统。**结果** 构建了以并发症名称、临床表现、病因、治疗措施等为本体层的造口相关并发症知识图谱,并建立了通过 Cypher 语言进行知识查询的智能问答系统。**结论** 本智能问答系统,可帮助肠造口患者快速获取经过循证的医疗信息,赋能肠造口患者并发症的科学自我管理,促进肠造口管理服务普惠化。

【关键词】 肠造口;并发症;知识图谱;问答系统;护理

doi:10.3969/j.issn.2097-1826.2024.11.003

【中图分类号】 R47-05 **【文献标识码】** A **【文章编号】** 2097-1826(2024)11-0009-04

Construction of a Question-Answering System for Enterostomy-Related Complications Based on Knowledge Map

CHEN Yanan¹, CHEN Ziting¹, XIE Yige¹, SHANG Haobin², WANG Zhiwen¹, ZHANG Xu¹ (1.School of Nursing, Peking University, Beijing 100191, China; 2.School of Electronics Engineering and Computer Science, Peking University)

Corresponding author: ZHANG Xu, Tel: 010-82805242

【Abstract】 Objective To establish an intelligent question-answering system based on the knowledge map of enterostomy-related complications. **Methods** (1) Construction of the knowledge map: Through integrating various knowledge sources such as guidelines related to enterostomy related complications, the ontology of the knowledge map of enterostomy-related complications was designed using the seven-step ontology construction approach. Subsequently, the Aho-Corasick automaton (AC automaton) multi-pattern matching algorithm was adopted to establish entity-relationship-entity triples of stoma complications, thereby constructing the knowledge map of stoma complications. (2) Construction of the intelligent question-answering system: An ontology of enterostomy-related complications natural language vocabulary and feature word vocabulary was established, and the questions were categorized by template matching. Then, the structured query language (SQL) matching statements appropriate for the Cypher (Cypher graph database query language) syntax were formed to accomplish the knowledge map query task, and ultimately, the intelligent question-answering system for enterostomy-related complications was established. **Results** A knowledge map of stoma-related complications was constructed, with entities such as complication names, clinical manifestations, etiologies, and treatment measures as the ontology layer. Additionally, an intelligent question-answering system for conducting knowledge queries through Cypher language was established. **Conclusions** This intelligent question-answering system enables patients with enterostomy to promptly obtain evidence-based medical information, empowers the scientific self-management of complications among patients with enterostomy, and promotes the universalization of enterostomy management services.

【Key words】 enterostomy; complications; knowledge map; question-answering system; nurse

【收稿日期】 2024-04-23 **【修回日期】** 2024-11-03

[Mil Nurs, 2024, 41(11): 9-12]

【基金项目】 中国博士后科学基金面上资助项目(2024M750136);
北京大学护理学科科研发展基金(TYZH2023001)

【作者简介】 陈雅楠, 本科在读, 电话: 010-82805242

【通信作者】 张旭, 电话: 010-82805242

我国结直肠癌的发病率和死亡率均居于世界首位^[1]。而肠造口术是根治结直肠癌的最常见治疗手

段。研究^[2-3]显示, 2.90%~81.10%的造口患者可能发生造口相关并发症, 常见造口相关并发症包括造口缺血、坏死、造口旁疝等^[4-5]。造口相关并发症的发生严重影响患者生活^[6]。近年来, 造口患者数量与日俱增, 造口门诊平均每年需接诊数千人次, 我国面临着专业造口治疗师短缺、专科化建设不足的瓶颈问题^[7], 极大限制了造口患者获取优质护理服务和造口相关健康教育的渠道和资源。智能问答系统可自动回复用户的个性化问题, 为远程造口并发症管理指导, 解决造口师不足的问题创造了可能。因此, 本研究为造口患者构建了基于肠造口相关并发症知识图谱的智能问答系统, 将造口相关并发症症状、预防措施等信息作为实体, 建立实体间的空间联结, 利用结构化知识图谱实现问题理解和答案推断, 智能回答患者个性化提问, 赋能患者科学自我管理, 促进肠造口管理服务普惠化。

1 资料与方法

1.1 资料来源 以《造口护理学》和《伤口造口失禁专科护理》为基础作为本体层结构指导, 标注肠造口相关并发症知识实体与实体关系。在万方、中国知网等以“肠造口并发症护理指南” OR “造口并发症管理”检索中文指南, 共获文献 29 篇; 在 PubMed、Web of Science、CINHAL 等数据库检索“Practice guideline” and “Enterostomy Complications”, 共获文献 23 篇; 此外在世界造口治疗师协会 (World Council of Enterostomal Therapists, WCET)、国际造口协会等相关网站进行系统检索, 进一步丰富数据层语料来源。检索时间为资源库建立至 2024 年 7 月 17 日。

1.2 肠造口相关并发症知识图谱的构建

1.2.1 知识图谱的本体 (schema) 层构建 全国造口治疗师规范化培训教材《造口护理学》和中华护理学会专科护士培训教材《伤口造口失禁专科护理》是国内公认的权威性护理学专著, 因此本研究选择以上教材进行本体构建; 采用斯坦福大学医学院开发的本体构建七步法^[8], 构建肠造口相关并发症知识图谱本体: (1) 确定本体的专业领域为肠造口相关并发症管理, 服务用户是肠造口患者, 目的是使肠造口相关并发症管理知识结构化并构建知识图谱, 实现查询及问答功能, 从而缓解造口治疗师短缺的问题; (2) 目前的肠造口相关并发症管理领域的本体结构尚缺乏, 无可复用的本体; (3) 确定本项目的重要术语, 包括肠造口相关并发症、病因 (腹壁伤口, 粪便污染源等)、临床表现 (疼痛、出血等) 及管理措施四个概念术语; (4) 明确管理措施的二级类目为预防 (定期更换造口袋等) 和干预 (就医、定期清洁造口等);

(5) 定义类的对象属性即病因-导致-肠造口相关并发症、肠造口相关并发症-具有-临床表现、管理措施-改善-肠造口相关并发症和病因; 确定管理措施中干预和预防的属性分面取值类型均为文本; (6) 以机械性损伤为例构建实例, 参考《成人肠造口护理标准》、《伤口造口失禁专科护理》及《造口护理学》, 以知识图谱三元组模式进行描述 (如图 1)。

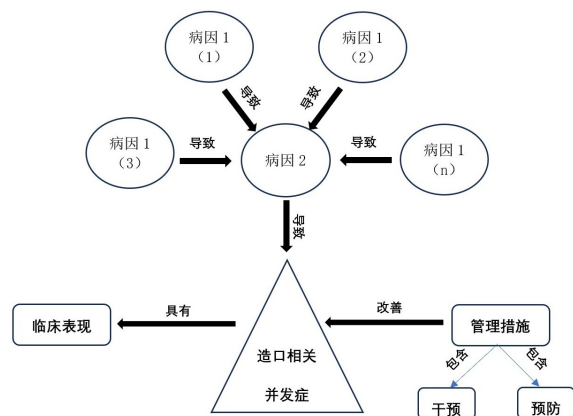


图 1 本体的语义关系

1.2.2 知识图谱的数据层构建 本研究基于设计的本体层, 使用人工筛选的方法从教材等知识源中抽取造口病因、临床表现等知识, 对管理措施等实体进行分类, 丰富实体内容。关系抽取采用人工标注方法, 知识源中进行实体及实体关系的标注, 以多种知识来源丰富实体及实体关系, 避免知识的缺失和遗漏, 保证标注的准确性。对于不同知识源的实体, 经常具有“多词一义”“缺字漏字”的情况, 则人工校对进行实体对齐, 实现医学知识融合, 以得到准确全面的实体关系三元组, 将其存储于 Neo4j 图数据库。

1.2.3 知识存储与查询 本研究采用 Neo4j 图数据库进行造口相关并发症知识图谱存储, 基于 Python 编程将实体关系三元组可视化, 生成造口相关并发症知识图谱; 数据层构建完成后将其整理成为“Id-实体类型-实体名称”的三元组表格和“Id-关系-Id”的三元组表格, 存储为 XLSX 文件并在微软 Excel 中转换为用于存储表格和电子表格信息的纯文本 (comma separated values, CSV) 文件格式; 最后利用 Neo4j 数据库用户手册中的语句将 CSV 文件导入到数据库中形成知识图谱。节点代表实体, 节点间的边表示实体间的语义关系。同时, Neo4j 数据库支持 CypherNeo4j 的声明式图形查询语言, 引入 Python 模块中的 Py2neo 模块实现 Python 语句对 Neo4j 数据库的可编程性, 然后通过 Python 实现对 Neo4j 数据库的连接、查询等操作, 实现造口相关并发症知识图谱的自然语言查询,

并可通过智能问答系统展示查询结果。

1.3 基于肠造口相关并发症知识图谱的智能问答系统构建 基于开发的知识图谱继续构建中文智能问答系统,构建的问答系统分为以下 5 个模块:(1) 用户交互界面;(2) 生物医学实体识别与问题分类;(3) 问题解析、模板匹配与 Cypher 查询语句构建;(4) 基于 Neo4j 数据库的答案查询与结果展示;(5) 用户反馈与系统评价。该结构如图 2 所示:以肠造口患者作为目标用户,将交互界面设计成通俗易懂的形式,便于患者使用。

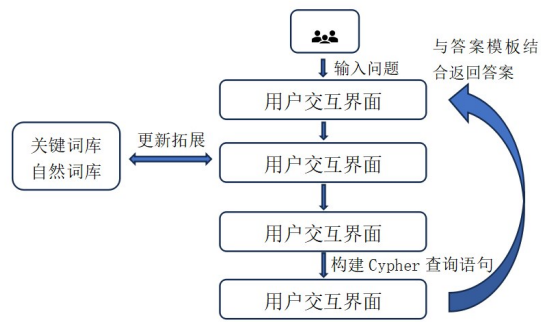


图 2 问答系统模块结构图

1.3.1 生物医学实体识别与问题分类 本研究基于上述知识图谱的节点构建了该问答系统的主要生物医学实体集合[disease]、[symptom]、[cause]、[prevent]、[improve],并设定了 4 种基本问题的类别:(1)已知症状筛查可能发生的并发症;(2)发生并发症的原因;(3)发生并发症的护理措施;(4)相关并发症/症状的预防措施,分别对应到后 4 个相应的实体集合中。基于语义分析,该系统定义了特征词、特征词路径、领域多模式匹配创建的工具库(领域 Ac-tree)、词典和问题疑问词,在识别到问题后对用户的自然语言进行分类与分析,从而得到主要生物医学实体和相应问题。由于该智能问答系统主要是面向造口患者,本研究在分类的基础上进一步构建了每一种基本问题关键词的自然语言词库,如表 1 所示,由此识别结果是一个形如[医学实体,问题关键词]的列表,例如,问题“造口水肿怎么预防?”,识别结果为[“造口水肿”,“预防”]。由此,该系统就能对用户所提问题进行有效的、有规律的归类,以便后续解析与模板匹配。

1.3.2 问题解析、模板匹配与 Cypher 查询语句构建 常见的用于问题理解的技术有基于模板匹配、基于检索模型及基于深度学习的模型方法。本研究应用了模板匹配的方法,将患者输入的问题根据自然语言词库进行分类并匹配其中的自然语言结果模板,即根据所得的[医学实体,问题关键词]构建在

Neo4j 中标准的 Cypher 查询语句。本研究使用的是 Neo4j 图形数据库的 Cypher 查询语言,遵循 SQL 语法,根据预先的[医学实体,问题关键词]分类与匹配的模板与知识图谱的关联映射可以推断出特定的关系,继而根据规则生成查询语句用于查询与已知实体有特定关系的相关实体名。基于本研究的知识图谱构建的 Cypher 语句模板如下:Match (a:dis-ease)-(r: QuestionName)-(b) where b.name = ‘EntityName’ return a.name,b.name,其中 QuestionName 和 EntityName 用相应的问题关键词和实体替换。

表 1 基于知识图谱的肠造口相关并发症问答系统基本问题关键词的自然语言词库

医学实体集合	问题关键词	自然语言词库
[Symptom]	临床表现	[症状,表征,现象,症候,表现]
[Cause]	病因	[原因,为什么,为啥,为何,怎么会,怎样会,怎么才会,怎么回事,咋样会,如何会,如何才会,导致,造成,产生]
[Improve]	干预	[缓解,改善,疗法,怎么办,怎么弄,怎么做,怎么治,咋办,咋治,咋弄,如何做]
[Prevent]	预防	[预防,防范,抵御,避免,免得,怎么才不,咋样才不,如何才能不]

2 结果

2.1 肠造口相关并发症知识图谱实体本体与实体关系 该知识图谱以病因、临床症状、管理措施等为一级知识框架,包含 17 种造口相关并发症相关的 3730 个实体,形成了 4860 个语义关系。本研究的主要语义关系,见表 2。

表 2 肠造口相关并发症知识图谱主要语义关系框架

起始实体	关系	指向实体
病因 1	导致	病因 2
病因 2	导致	造口相关并发症
管理措施	改善	造口相关并发症
管理措施	改善	病因 2
造口相关并发症	具有	临床表现
管理措施	包含	干预措施
管理措施	包含	预防措施

2.2 肠造口相关并发症知识图谱与智能问答系统可视化呈现与应用 造口相关并发症知识图谱以肠造口相关并发症为核心,围绕病因、症状、管理措施、临床表现等 4 类知识实体逐一展开,可以通过鼠标指向查看实体内容或双击打开知识实体的详细内容,每一条知识实体的阐述内容均是多种来源知识聚合的结果。本研究构建的造口相关并发症知识图谱存储在 Neo4j 数据库,通过 Web 端将数据库与互联网链

接,形成了基于 Web 端的肠造口相关并发症知识图谱智能问答系统(图 3),用户在该界面中输入相关知识实体的自然语言,系统通过自然语言解析,从 Neo4j 数据库自动匹配相关数据,并将结果返回至该页面,以可视化图谱和详细的文字描述呈现应答内容。

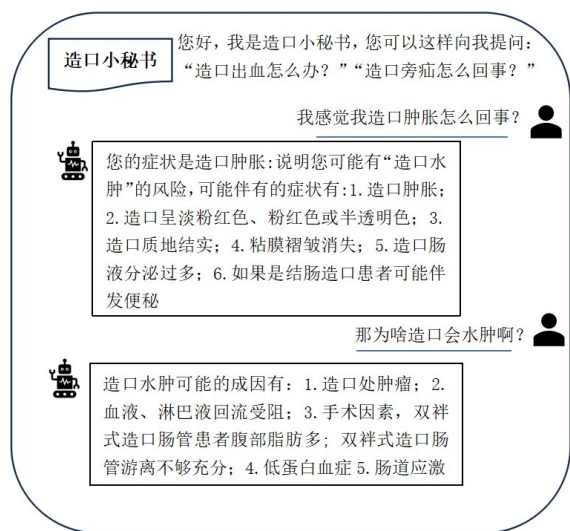


图 3 Web 端问答系统展示

3 讨论

3.1 造口相关并发症知识图谱和问答系统可增强患者健康管理能力 由于医学知识的专业性,用户难以在互联网快速获取所需的医疗知识^[9]。同时造口患者大多自我照护能力和自我掌控水平较低^[10],因此向患者及其照护者提供经过循证的医疗知识具有重要的现实意义。基于知识图谱的智能问答系统,能够充分利用知识图谱中的结构化数据,为用户反馈简洁、精确的答案^[11],但相较于国外,国内医学领域智能问答系统的研究尚在起步阶段^[12],并且大多问答系统面向的是全科疾病的轻量级内容问答,很少根据造口相关并发症提供专业的深度知识问答服务。本问答系统涵盖症状、病因、管理措施相关循证知识,可针对常见造口相关并发症准确解答。此外,基于自然语境引入的自然问答词库,将专业内容通俗化与简单化,提高了用户交互性与掌握程度。通过移动设备进行咨询与问答,为造口患者提供了准确获取造口相关并发症知识与健康教育的渠道^[13],更具针对性、普适性。本智能问答系统可以小程序的形式在微信等移动平台运行,有助于实现健康教育的个性化^[14],提高患者自我健康管理的能力。

3.2 造口相关并发症知识图谱和问答系统有较为广泛的应用前景 在卫生信息化进程不断深入、人工智能技术井喷背景下^[15],基于医学知识图谱开发

智能问答系统可帮助造口患者自查自诊,缓解造口师短缺对优质护理服务的制约,为更多的造口患者提供随时在线的“虚拟造口师”,进一步推进“互联网+医疗”的建设^[16]。本问答系统可以结合拓展应用范围,嵌入线上平台,为更大范围、更多途径地向公众提供造口并发症自我筛查相关知识奠定了理论和技术基础。本研究未来拟增加更多专家咨询与用户体验的反馈环节,同时将深度学习技术融入到系统中,使系统更加智能化,以提高其对自然语言问题理解的准确性。此外,为了更好地与其它医疗相关应用和线上平台结合,我们将持续进行用户端的优化,完善其实际应用效果,以提高用户使用的舒适感。

【参考文献】

- [1] 郑荣寿,张思维,孙可欣,等.2016 年中国恶性肿瘤流行情况分析[J].中华肿瘤杂志,2023,45(3):212-220.
- [2] 中国造口管理协作组,中华护理学会伤口造口失禁护理专业委员会,中国医师协会外科医师分会结直肠医师专业委员会,等.肠造口并发症的分型与分级标准(2023 版)[J].中华胃肠外科杂志,2023,26(10):915-921.
- [3] CHABAL L O,PRENTICE J L,AYELLO E A.Practice implications from the WCET® International ostomy guideline 2020[J].Adv Skin Wound Care,2021,34(6):293-300.
- [4] 王玲,胡爱玲.伤口造口失禁专科护理[M].北京:人民卫生出版社,2018:53-87.
- [5] 丁炎明.造口护理学[M].北京:人民卫生出版社,2017:175-238.
- [6] 闫金英,徐征,赵杰,等.肠造口病人早期造口并发症的现状影响因素研究[J].全科护理,2021,19(24):3447-3449.
- [7] 顾佳妮,刘萍,姜丽萍.家庭参与式护理模式在成人肠造口患者中的应用效果[J].军事护理,2023,41(11):57-61.
- [8] 甘健侯,姜跃,夏幼明.本体方法及其应用[M].北京:科学出版社,2011:115-137.
- [9] SWIRE-THOMPSON B,LAZER D.Public Health and online misinformation:challenges and recommendations[J].Annu Rev Public Health,2020,41(1):433-451.
- [10] MINIOTTI M,BASSINO S,FANCHINI L,et al.Supportive care needs, quality of life and psychological morbidity of advanced colorectal cancer patients[J].Ann R Coll Surg Engl,2019(43):501-508.
- [11] 李俊卓,管红英,闫英杰,等.儿科疾病及保健知识问答系统的构建[J].中文信息学报,2022,36(1):127-134.
- [12] 张芳芳,马敬东,王小贤,等.国外医学领域自动问答系统研究现状及启示[J].医学信息学杂志,2017,38(3):2-6,25.
- [13] 孙斌,常开志,李树涛.面向医疗咨询的复杂问句意图智能理解[J].中文信息学报,2023,37(1):112-120.
- [14] 盛瑶雪,邹志辉,付亭亭,等.移动健康教育平台在带 T 管出院患者居家护理中的应用[J].解放军护理杂志,2018,35(3):73-76.
- [15] 韩国元,冷雪忠.新医改十年:中国医疗卫生信息化多元政策主体协同治理研究[J].北京航空航天大学学报:社会科学版,2024,37(4):113-126.
- [16] 范媛媛,李忠民.中文医学知识图谱研究及应用进展[J].计算机科学与探索,2022,16(10):2219-2233.

(本文编辑:沈园园)