

# 临床实践指南实施性促进研究之二： 基于非肌层浸润性膀胱癌指南的知识图谱 框架设计



王永博<sup>1#</sup>, 高 旷<sup>2#</sup>, 李绪辉<sup>1</sup>, 黄 桥<sup>1</sup>, 郭 静<sup>3</sup>, 司宜蓓<sup>4</sup>, 陈沐坤<sup>2</sup>, 阎思宇<sup>1</sup>,  
胡文斌<sup>2</sup>, 靳英辉<sup>1</sup>

1. 武汉大学中南医院循证与转化医学中心 (武汉 430071)
2. 武汉大学计算机学院 (武汉 430072)
3. 南京中医药大学附属医院针灸康复科 (南京 210029)
4. 武汉大学第二临床学院 (武汉 430071)

**【摘要】** 指南的转化和应用是医学科技成果转化的关键环节。但是, 目前指南在临床决策和实践可操作性方面仍存在较多不足。知识图谱技术的引入为指南智能化提供了坚实的基础, 为解决指南实施问题提供了良好的手段。非肌层浸润性膀胱癌(non-muscle invasive bladder cancer, NMIBC)是泌尿系统常见的恶性肿瘤, 随着新的诊疗手段不断涌现, NMIBC 临床指南存在查阅费时、更新频繁等问题。鉴于临床医生迫切需要学习和使用指南知识、梳理指南更新内容, 本研究基于 NMIBC 疾病特征和临床指南特点, 提出并构建 NMIBC 临床指南知识图谱框架, 包括解析指南内容范围、梳理指南内容模块、归纳指南知识结构等, 参考 OMAHA Schema, 根据二级概念层及概念实例设计概念结构表, 设计并调整概念间的关系结构, 改进并完善实体与实体间的关系, 并请专家对知识图谱的科学性和方法有效性进行审核与评估。基于 NMIBC 指南的知识图谱框架设计为指南数字化、智能化提供了基础, 有利于推动指南的实施与传播。

**【关键词】** 非肌层浸润性膀胱癌; 临床指南; 知识图谱; Neo4j

Research on promotion of implementation of clinical practice guidelines (II):  
framework design of knowledge graph construction based on guidelines for  
non-muscle invasive bladder cancer

Yong-Bo WANG<sup>1#</sup>, Kuang GAO<sup>2#</sup>, Xu-Hui LI<sup>1</sup>, Qiao HUANG<sup>1</sup>, Jing GUO<sup>3</sup>, Yi-Bei SI<sup>4</sup>, Mu-Kun CHEN<sup>2</sup>,  
Si-Yu YAN<sup>1</sup>, Wen-Bin HU<sup>2</sup>, Ying-Hui JIN<sup>1</sup>

1. Center for Evidence-Based and Translational Medicine, Zhongnan Hospital of Wuhan University,  
Wuhan 430071, China

2. School of Computer Science, Wuhan University, Wuhan 430072, China

3. Acupuncture and Rehabilitation Department, The Affiliated Hospital of Nanjing University of

DOI: 10.12173/j.issn.1004-5511.2020111058

# 共同第一作者

基金项目: 国家重点研发计划“科技助力经济 2020”重点专项(SQ2020YFF0426494); 国家自然科学基金项目(82174230); 武汉大学中南医院 2020 年度转化医学及交叉学科研究联合基金项目(ZNJC202016)

通信作者: 胡文斌, 博士, 教授, 博士研究生导师, E-mail: hwb@whu.edu.cn

靳英辉, 博士, 副教授, 硕士研究生导师, E-mail: jinyinghui0301@163.com

Chinese Medicine, Nanjing 210029, China

4. The Second Clinical College, Wuhan University, Wuhan 430071, China

\*Co-first author: Yong-Bo WANG and Kuang GAO

Corresponding author: Wen-Bin HU, E-mail: hwb@whu.edu.cn; Ying-Hui JIN, E-mail: jinyinghui0301@163.com

**【Abstract】** The transformation and application of the guidelines is a crucial link in the transformation of medical science achievements and technology. However, there are still many problems with the operability of the current guidelines in clinical decision-making and practice. The knowledge graph has provided a solid foundation for the intelligentization of the guidelines and a good method for solving the problems of the implementation of the guidelines. Non-muscle invasive bladder cancer (NMIBC) is a common malignant tumor of the urinary system, and new diagnosis and treatment methods for NMIBC are emerging. In view of the time-consuming and frequent update of clinical guidelines for NMIBC, as well as the urgent need for clinicians to learn, use, and sort out the updated content of the guidelines, this study is based on the disease characteristics of NMIBC and the clinical guidelines to propose and construct a knowledge graph framework for clinical guidelines of NMIBC, which mainly starts from analyzing the scope of the guideline, combing the guideline content modules, summarizing the guideline knowledge structure. Then, we referred to the OMAHA Schema, and designed the concept structure table according to the secondary concept layer and concept instances. Thereafter, we designed and adjusted the relationship structure between concepts, improved the relationship between entities, and finally reviewed and evaluated the scientific and method validity of the knowledge graph through experts. The design of the knowledge graph construction framework based on the guideline for NMIBC proposed in this paper provides a basis for the digitization and intelligence of the guideline, which is conducive to the implementation and dissemination of the guideline.

**【Keywords】** Non-muscle invasive bladder cancer; Clinical guidelines; Knowledge graph; Neo4j

临床指南是基于系统评价证据, 平衡不同干预措施的利弊后形成的旨在为医生和患者提供符合实际情况的推荐意见<sup>[1]</sup>。临床指南的内容不仅整合了该领域权威专家的临床指导意见, 而且涵盖了基本的临床相关理论知识和科学严谨的临床诊疗建议。作为衔接循证医学和临床实践的纽带, 临床指南在规范医疗行为、提高医疗质量、降低医疗成本、减少医疗纠纷等方面发挥着重要作用。指南的转化和应用是医学科技成果转化的关键环节。但是, 有研究指出目前指南的传播仍然局限于文本形式, 导致临床指南的内容不能被实时、准确地查览, 严重阻碍了指南在临床决策和实践中的可操作性, 使其难以发挥真正的价值<sup>[2-4]</sup>。

“人工智能+医疗”是指人工智能通过机器学习、深度学习、自然语言处理等技术, 利用计

算机算法从数据中获取信息, 实现辅助诊断、疾病分诊、疗法选择、风险预测等一系列功能<sup>[5]</sup>。随着人工智能的快速发展, 知识图谱已经成为知识服务领域的研究热点<sup>[6]</sup>。Google 于 2012 年首次引入了知识图谱的概念, 并将其应用于搜索引擎, 以提高搜索效率。知识图谱以结构化的形式描述了客观世界中的概念、实体及其之间的关系, 将互联网的信息表达以更接近人类认知世界的形式呈现, 提供了一种更好的组织、管理和理解互联网海量信息的能力<sup>[7]</sup>。目前, 知识图谱在医学领域拥有着广泛的应用和发展前景, 如疾病风险评估、智能咨询诊疗、医疗质量控制和医疗知识问答等<sup>[8]</sup>。构建基于临床指南的医学知识图谱对辅助临床决策有着重要研究价值。

膀胱癌是全球第十大常见肿瘤<sup>[9]</sup>。全球范围

内,男性年龄标化发病率为9.5/10万人,女性2.4/10万人;男性年龄标化死亡率为3.3/10万人,女性0.86/10万人<sup>[9]</sup>。膀胱癌可分为非肌层浸润性膀胱癌(non-muscle invasive bladder cancer, NMIBC)和肌层浸润性膀胱癌。约75%的膀胱癌患者在初次诊断时为NMIBC<sup>[10]</sup>。近年来,针对NMIBC新的诊断、治疗方式不断涌现,为此,世界各国和地区不断制订、更新NMIBC临床指南,以指导医生的临床实践。我国也积极推进相关指南制订工作。本课题组于2018年发表了《中国非肌层浸润性膀胱癌治疗与监测循证临床实践指南(2018年标准版)》<sup>[11]</sup>。

目前鲜有研究围绕NMIBC临床指南知识图谱设计与应用展开探讨。本文提出并构建NMIBC临床指南知识图谱框架,旨在为指南数字化、智能化提供基础,以推动指南的实施与传播。

## 1 知识图谱框架设计

### 1.1 知识图谱框架蓝图

基于NMIBC疾病特征、诊疗现状和指南特点以及临床应用常见问题,明确NMIBC临床指南的知识图谱构建目标,设计符合专业认知与实际需要的NMIBC临床指南知识图谱,图1为NMIBC知识图谱构建框架蓝图。

### 1.2 指南内容结构化

本研究以《中国非肌层浸润性膀胱癌治疗与监测循证临床实践指南(2018年标准版)》<sup>[11]</sup>为例。该指南共包含8项主题,分别是:NMIBC手术治疗、化疗、免疫治疗、联合治疗、NMIBC原位癌治疗、NMIBC患者行根治性膀胱切除术、NMIBC复发治疗、随访及监测。通过对指南内容进行梳理,明确该指南涵盖的基本诊治过程,如根据患者基本信息、疾病特征、诊断,提供相应的治疗、预后预测及随访流程。据此,对指南的内容范围进行解析,梳理归纳指南内容和知识结构,结合专家意见设置NMIBC中各概念间的层级关系,对指南内容进行结构化组织(图2)。

### 1.3 概念层(本体层)设计

#### 1.3.1 概念结构设计

知识图谱主要有自顶向下、自底向上及综合法三种构建方式,本研究采用综合法。该方法首先需定义大量重要概念,然后将它们分别进行恰当的归纳和演绎,并与一些中级概念关联起来。

首先,为知识图谱定义概念层与数据层模式,并将实体与关系加入知识库。概念是指具有同种特性的实体构成的集合。本体是结构化知识库的概念模板,通过概念层设计而形成的知识库不仅层次结构较强,且冗余程度较小。概念层是数据

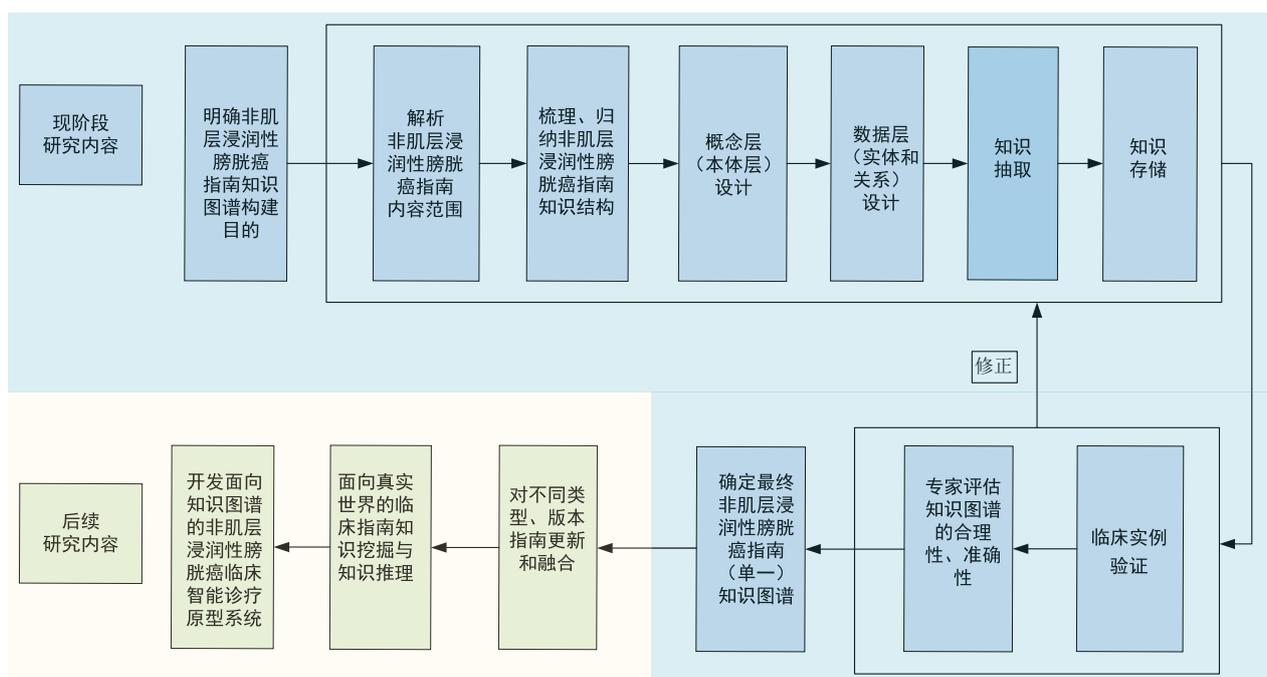


图1 NMIBC指南知识图谱构建的框架蓝图

Figure 1. The framework blueprint for the construction of the NMIBC guideline knowledge graph

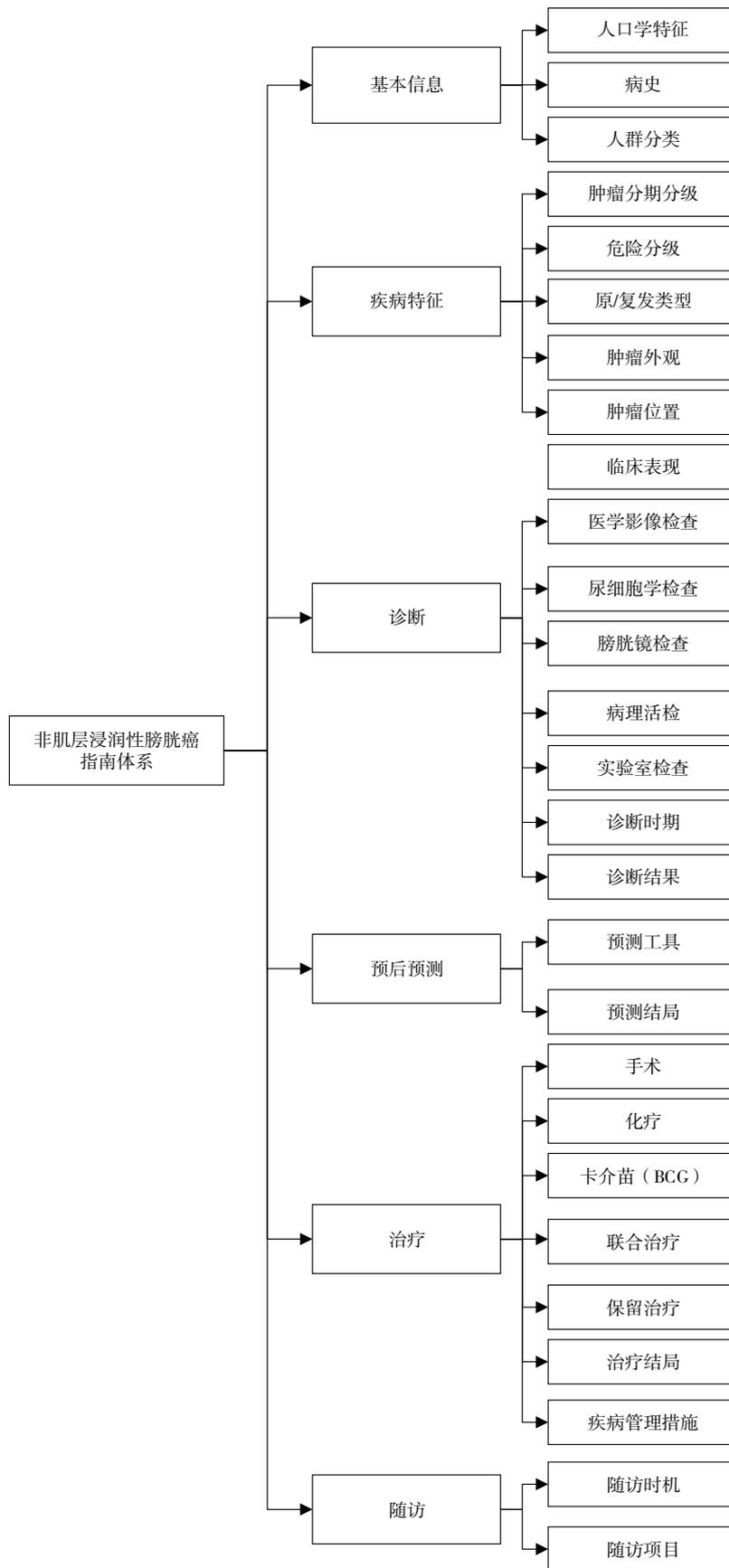


图2 NMIBC指南体系

Figure 2. NMIBC guideline system

的模式，是对数据层的提炼。数据层主要由一系列的事实组成，而知识将以事实为单位进行存储，即数据层是具体的数据。概念层构建在数据层之上，是知识图谱的核心，需依据其来管理和组织数据层。

本研究参照 OMAHA Schema（当前版本更新于 2021 年 8 月 20 日）进行概念层提取<sup>[12]</sup>。OMAHA Schema 旨在为中文医学领域的知识图谱构建、数据挖掘、语义分析等提供可参考和可扩展的数据定义和描述规范。在 Schema 的构建过程中，充分研究和参考了国内外成熟的知识图谱 Schema（UMLS 语义网络、Schema.org、cnSchema 等），同时也考虑了中文医学知识的特点，是一

套符合中文环境的医学知识图谱。NMIBC 指南体系的大部分内容都符合 OMAHA Schema 规范，对临床指南中 OMAHA Schema 未包含的概念层，本研究通过反复讨论进行补充，整理出的 NMIBC 本体二级概念结构，如表 1 所示。

### 1.3.2 实体语义关系设计

为实现指南结构的整合，其中一个重要的步骤是使这些概念层在语义上形成关联，以构成一个完整的指南结构体。本研究团队根据 OMAHA Schema 设计了一张以 NMIBC 知识图谱为使用对象的实体语义关系表（表 2，图 3），不同疾病知识图谱的概念关系可在此表的基础上进行设计并使用，也可对特定关系进行补充。

表1 NMIBC指南本体概念结构（部分）

Table 1. Conceptual structure for NMIBC guideline (part)

一级概念层	一级概念定义	二级概念层	实例
文献资料	通过一定的方法和手段、运用一定的意义表达和记录体系记录在一定载体的有历史价值和研究价值的知识	临床实践指南	《中国非肌层浸润性膀胱癌治疗与监测循证临床实践指南（2018年标准版）》，EAU Guidelines on Non-muscle-invasive Bladder Cancer (TaT1 and CIS)
临床问题	在临床循证医学实践中提出的临床需要解决的问题	危险因素/病因 诊断 治疗 预后	临床问题1 临床问题2 临床问题3 临床问题4
推荐意见	临床指南的临床问题所对应的推荐意见		推荐意见1
人	人物	人群	NMIBC患者，中高危NMIBC患者，不适合行经尿道膀胱肿瘤切除术（transurethral resection of bladder tumor, TURBT）的患者，怀疑有多发性肿瘤、原位癌（carcinoma in situ, CIS）的患者
疾病	在一定的病因作用下，机体内稳态调节紊乱而导致的生命活动障碍		高危肿瘤、T1期肿瘤、G3/高级别肿瘤（CIS除外）、活动性肺结核
临床所见	当前状况下观察到的表现，或者是患者自身经历并作为主诉的表现	体征 症状	淋巴结肿大、消瘦、贫血 TURBT术中疑似发生膀胱穿孔，TURBT术后严重肉眼血尿，BCG治疗的禁忌证，TURBT术后泌尿道感染，尿路症状（urinary tract symptoms），尿频、尿急、尿痛、排尿困难、尿路刺激
事件*	人为操作与干预的临床过程	操作治疗方案 药物治疗方案 诊断方案	操作治疗方案1 药物治疗方案1 诊断方案1
人体形态与结构	表示正常或异常的人体组织器官的位置、形态、结构、功能及其发展规律	正常形态结构 异常形态结构	基底部分膀胱壁 内镜可见肿瘤，1 cm以内的肿瘤，较大肿瘤

续表 1

一级概念层	一级概念定义	二级概念层	实例
操作	表示为诊断、预防、治疗疾病或改善身体机能所采取的操作、方法或技术		TURBT、即刻膀胱灌注化疗、膀胱灌注化疗、热灌注疗法、膀胱内BCG灌注治疗、经尿道前列腺切除术
分子活性	表示分子水平上的生理功能		基因组、蛋白质组
基因突变	表示基因组DNA分子发生的突然的、可遗传的变异现象		FGFR3基因突变
物理实体	表示视觉或触觉可感知的物体	医疗器械	荧光成像引导、肾脏超声（renal ultrasound）、膀胱超声（bladder ultrasound）、CT检查、计算机断层扫描静脉尿路造影（computed tomography-intravenous urography、CT-IVU）、尿路造影（CT urography）、MRI、柔性膀胱镜（flexible cystoscope）、一体双通道膀胱导管镜
药品	表示由生产厂家生产的药物制剂，名称通常包括物质、强度和形式等		吉西他滨、吡柔比星、羟基喜树碱、丝裂霉素、多柔比星、表柔比星、BCG
限定词	表示特指或类指等限定作用的词	严重程度	严重
		优先级别	优先使用
		治疗环境	设备、操作人员条件允许
		分期状态	Ta期、T1期、CIS、TaG1、G3/高级别、低级别、TaLG/G1、TaLG/G1-2、无法明确分期
		意图	病理性诊断，减少尿液排泄来维持药物浓度，提高膀胱内灌注化疗疗效
		剂型	片剂、注射剂
		医学学科	泌尿外科
		单位	克
		就诊类型	门诊、住院
		指南状态	现行
		指南类型	治疗与监测循证临床实践指南
		操作方法	切除
		时间状态	TURBT术后24 h、0.5~2 h、1~3年、即刻
		治疗阶段	初始治疗（initial treatment）
		用药目的	预防、治疗
		病理过程	黏膜异常（mucosal abnormalities）
		给药途径	膀胱灌注
		药物处方分类	处方药、非处方药
		评价结果	正常、异常
		语种	英文、中文
		身体状态	静息

注：\*事件的概念由作者自定义

表2 NMIBC指南本体设计中实体语义关系  
(部分)

Table 2. Entity semantic relation in NMIBC guideline ontology design (part)

起始实体类别	关系名	指向实体类别
临床问题	包含于	文献资料
临床问题	推荐意见	推荐意见
临床问题	相关人群	人群
临床问题	干预方法	操作
临床问题	对比方法	操作
推荐意见	涉及	操作
推荐意见	涉及	操作治疗方案
推荐意见	涉及	观测操作
推荐意见	相关人群	人群
操作治疗方案	操作	操作
操作治疗方案	操作部位	人体形态与结构
操作治疗方案	操作意图	意图
操作治疗方案	持续时间	时间状态
操作治疗方案	禁忌证	临床所见
操作治疗方案	描述	文本
操作治疗方案	前提条件	治疗环境
操作治疗方案	适应证	临床所见
操作治疗方案	适应证	异常形态结构
操作治疗方案	适用人群	人群
操作治疗方案	术后治疗	操作
操作治疗方案	术中操作	操作
操作治疗方案	选择用药	药物
操作治疗方案	引导操作	操作
操作治疗方案	最佳操作时间	时间状态
临床所见	发生于……期间	文本
临床所见	临床表现	临床所见
临床所见	症状	临床所见
操作	不良反应	临床所见
操作	操作意图	意图
操作	持续时间	时间状态
操作	辅助性物质	药物
操作	禁忌证	临床所见
操作	适应证	临床所见
操作	前置操作	操作
人群	包含	人群

### 1.4 数据层 ( 实体和关系 ) 设计

实体是指具有可区别性且独立存在的某种事物。实体是知识图谱中最基本的元素,不同的实体间存在不同的关系。关系用于刻画实体和实体之间的联系。

构建指南数据层的目的是从不同格式的临床指南中提取医学知识,并将医学知识转换为可被计算机处理的数据格式。本研究基于三元组数据模型构建数据层,揭示指南内各类医学实体间的关联,并通过若干个组合的临床指南知识点,形成临床指南知识网络。

如表 3 所示,对于临床问题 1 “NMIBC 患者行 TURBT 的适应证”,推荐意见为“对怀疑为 NMIBC 的患者,推荐 TURBT 作为诊断和初始治疗的方法”。首先,根据 PICO 原则对该问题进行拆分,即人群为“NMIBC 患者”、干预方法为“TURBT”,另外添加问题的主题为“手术治疗”。其次,提取实例“临床问题 1”与“推荐意见 1”,两者分属“临床问题”和“推荐意见”概念层,概念“临床问题”与“推荐意见”的关系为“推荐意见”;提取实例“NMIBC 患者”为人群,概念“推荐意见”与“NMIBC 患者”的关系为“相关人群”;提取“TURBT”为操作,概念“推荐意见”与“TURBT”的关系为“涉及”;提取实例“NMIBC 患者”与“TURBT”的关系为“适应证”;提取“推荐”为文本,概念“推荐意见”与“推荐”的关系为“倾向”。临床问题 1 的主题为“手术治疗”,提取“临床问题 1”与“手术治疗”,两者关系为“主题”。问题 2—4 的数据层 ( 实体和关系 ) 设计如表 4 所示。

### 1.5 知识抽取

医学临床指南中含有大量医学术语,临床指南的知识抽取就是从临床指南中抽取与概念层设计相匹配的实体和关系,并将其整理成实体库和三元组关系库。针对 NMIBC 临床指南用语相对规范统一、内容结构较为明晰、句法搭配相对固定等特点,本研究选择人工抽取的方式。后续系列文章中会涉及到不同类型、不同版本指南的更新和融合,以及真实数据的挖掘过程,拟采用远程监督算法进行抽取。远程监督算法是一种半监督算法,可基于一个标注好的小型知识图谱,给外部文档库中的句子标注关系标签,以实体识别为基础完成关系抽取,远程监督关系抽取类似于

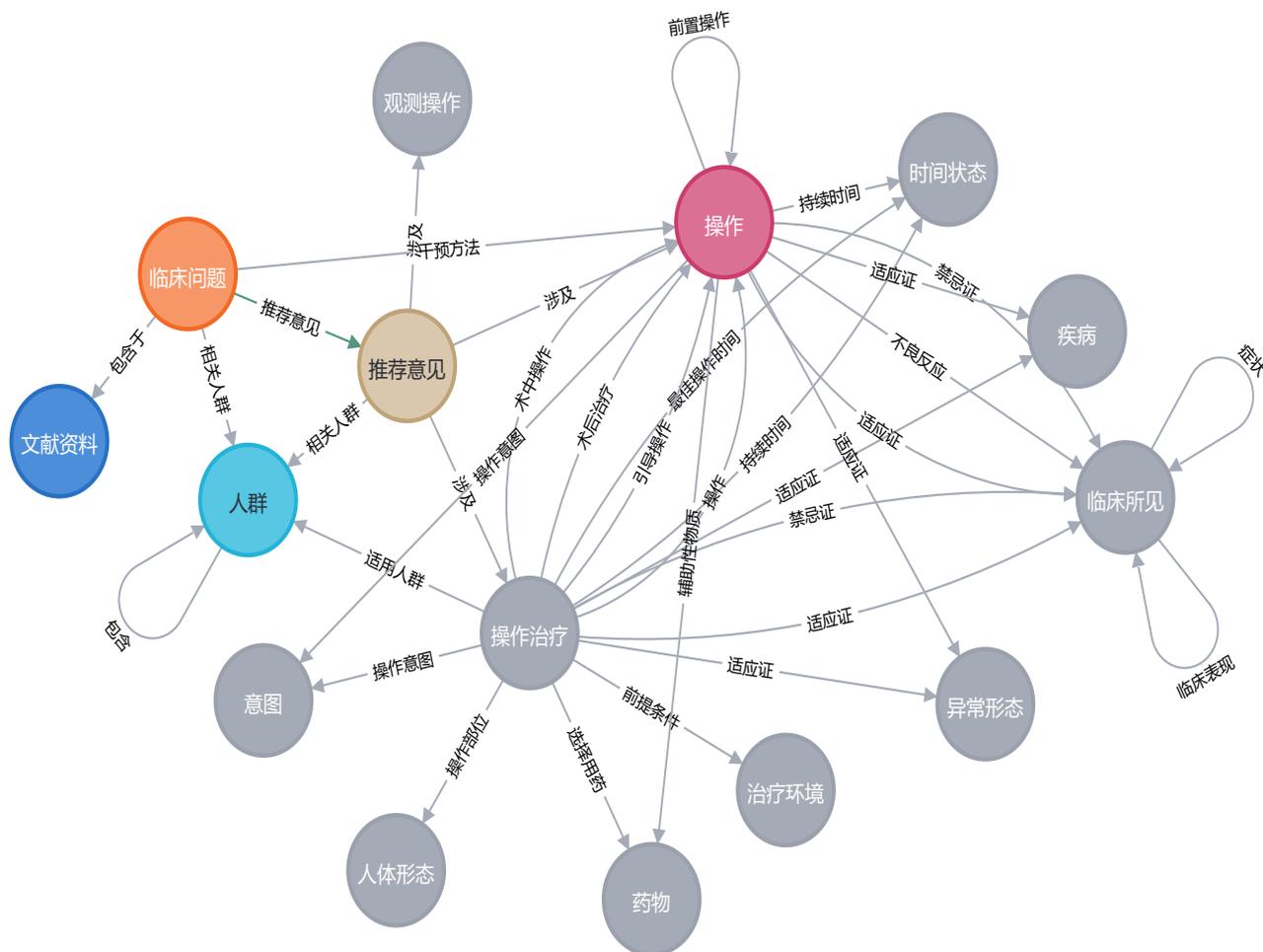


图3 NMIBC指南本体设计中实体语义关系图 (部分)

Figure 3. Entity semantic relation diagram in NMIBC guideline ontology design (part)

表3 指南部分问题及推荐意见

Table 3. Several questions and recommendations of the guideline

指南问题序号	指南问题	推荐意见序号	推荐意见
临床问题1	NMIBC患者行TURBT的适应证?	推荐意见1	对怀疑为NMIBC的患者, 推荐TURBT作为诊断和初始治疗的方法
临床问题2	术后即刻膀胱灌注化疗的禁忌证?	推荐意见2	术中发生膀胱穿孔或术后严重肉眼血尿的患者不宜行即刻膀胱灌注化疗
临床问题3	低危患者术后是否只需行即刻单次膀胱灌注化疗?	推荐意见3	对于低危NMIBC患者, 术后可仅行即刻单次膀胱灌注
临床问题4	对NMIBC患者, BCG诱导加维持灌注治疗是否优于单纯BCG诱导灌注治疗?	推荐意见4	推荐BCG诱导加维持灌注治疗

表4 三元组关系示例  
Table 4. Examples of triplet relation

指南问题	三元组指向内容	实体1	实体1所属概念	实体间关系	实体2	实体2所属概念	
临床问题1	临床问题三元组	临床问题1	临床问题	包含于	NMIBC治疗与监测 临床实践指南	文献资料	
		临床问题1	临床问题	相关人群	NMIBC患者	人群	
		临床问题1	临床问题	干预方法	TURBT	操作	
	临床问题到推荐意见	推荐意见三元组	临床问题1	临床问题	主题	手术治疗	文本
			临床问题1	临床问题	推荐意见	推荐意见1 <sup>a</sup>	推荐意见
			推荐意见1 <sup>a</sup>	推荐意见	相关人群	NMIBC患者	人群
			推荐意见1 <sup>a</sup>	推荐意见	涉及	TURBT	操作
	临床问题2	临床问题三元组	TURBT	操作	倾向	推荐	文本
			TURBT	操作	适应证	NMIBC	疾病
			临床问题2	临床问题	相关人群	NMIBC患者	人群
临床问题2	临床问题三元组	临床问题2	临床问题	干预方法	膀胱灌注化疗	操作	
		临床问题2	临床问题	主题	化疗	文本	
		临床问题2	临床问题	推荐意见	推荐意见2 <sup>b</sup>	推荐意见	
	临床问题到推荐意见	推荐意见三元组	临床问题2	临床问题	相关人群	NMIBC患者	人群
			推荐意见2 <sup>b</sup>	推荐意见	涉及	即刻膀胱灌注化疗	操作
			推荐意见2 <sup>b</sup>	推荐意见	禁忌证	TURBT术中疑似发生膀胱穿孔	临床所见
			即刻膀胱灌注化疗	操作	禁忌证	TURBT术后严重肉眼血尿	临床所见
	临床问题3	临床问题三元组	即刻膀胱灌注化疗	操作	禁忌证	TURBT术后严重肉	临床所见
			临床问题3	临床问题	相关人群	NMIBC患者	人群
			临床问题3	临床问题	干预方法	膀胱灌注化疗	操作
临床问题到推荐意见		推荐意见三元组	临床问题3	临床问题	主题	化疗	文本
			临床问题3	临床问题	推荐意见	推荐意见3 <sup>c</sup>	推荐意见
			推荐意见3 <sup>c</sup>	推荐意见	相关人群	低危NMIBC患者	人群
			推荐意见3 <sup>c</sup>	推荐意见	涉及	操作治疗方案3	操作治疗方案
临床问题4		临床问题三元组	操作治疗方案3	操作治疗方案	术中操作	TURBT	操作
			操作治疗方案3	操作治疗方案	术后治疗	即刻膀胱灌注化疗	操作
			临床问题4	临床问题	相关人群	NMIBC患者	人群
临床问题4	临床问题三元组	临床问题4	临床问题	干预方法	BCG诱导加维持灌 注化疗	操作	
		临床问题4	临床问题	对比方法	BCG诱导灌注治疗	操作	
		临床问题4	临床问题	主题	免疫治疗	文本	
	临床问题到推荐意见	推荐意见三元组	临床问题4	临床问题	推荐意见	推荐意见4 <sup>d</sup>	推荐意见
			推荐意见4 <sup>d</sup>	推荐意见	相关人群	NMIBC患者	人群
	临床问题4	临床问题三元组	推荐意见4 <sup>d</sup>	推荐意见	涉及	BCG诱导加维持灌 注化疗	操作
			推荐意见4 <sup>d</sup>	推荐意见	涉及	BCG诱导加维持灌 注化疗	操作

注：<sup>a</sup>证据级别为1A，推荐强度为强推荐；<sup>b</sup>证据级别为4，推荐强度为强推荐；<sup>c</sup>证据级别为4，推荐强度为强推荐；<sup>d</sup>证据级别为1A，推荐强度为强推荐；证据级别和推荐强度为相应实体的属性

多实例问题 (multiple instance problem), 借助多实例学习和远程监督算法可以快速完成知识的提取和展示。

针对临床问题 1—4 的问题和推荐意见, 实体抽取旨在获取如表 5 所示的结果。关系抽取参照表 2 和表 4 进行。若实体间的关系需要补充, 则通过专家讨论协商。

### 1.6 知识存储及生成知识图谱

明确了所有实体以及实体间的关系后, 需设计数据的物理模型对知识进行存储, 形成知识图谱。与传统的关系型数据库不同, 图形数据库基于图理论, 以非结构化的方式存储关联的数据, 将图形抽象为节点、边等基本元素, 其中一个节点代表一个实体, 节点间的边代表实体间的关系, 在关联表示方面具有高效的处理能力。为表示多样化、复杂化的医学实体关系, 本研究组选用 Neo4j 图数据库作为存储体系。基于抽取的三元组数据模型, 设计并导入医学节点和关系节点, 考虑到疾病临床指南整理的数据为中等规模以及数据可实时插入, 以 CSV 文件为主对数据进行存储。

将实体与关系整理成 CSV 格式的文件, 导入 Neo4j desktop 1.4.9 数据库, 输入 cypher 语句, 生成知识图谱:

```
LOAD CSV WITH HEADERS FROM "file:///
Neo4j_Load.csv" AS line
WITH line WHERE line.entityID is not Null and
line.valueId is not Null
match (from:entity{id:line.entityID}),
```

```
(to:entity{id:line.valueId})
```

```
merge (from)-[r:rel{property:line.property}]->(to)
```

其中, entityID 为起始实体 ID, valueID 为指向实体 ID, entity 为起始实体, value 为指向实体。

输入 cypher 语句:

```
MATCH (n) RETURN n
```

可得到可视化结果, 即 NMIBC 指南知识图谱全貌, 如图 4A 所示。

若想获取所有和“临床问题 1”相连接的实体及相应的关系, 则输入 cypher 语句:

```
match p=(临床问题 {name:'临床问题 1'})--()
return p
```

4 个临床问题相连接的实体及相应的关系如图 4B—E 所示。其中推荐意见采用绿色圆形表示, 其颜色深浅和节点圆形的大小分别表示推荐意见不同的证据等级和推荐强度, 使知识图谱节点展示更清晰直观, 查询或搜索时更加方便快捷。

## 2 临床实例验证

NMIBC 知识图谱可实现搜索与推荐功能, 应用于临床辅助决策。随着图数据规模的日益增大, 大图上的子图搜索问题变得极为重要<sup>[13]</sup>。以临床问题 1—4 为例, 通过后台代码演示数据库检索过程, 以验证检索的准确性。

### 2.1 查询某操作的适应证

对于临床问题 1: “TURBT 的适应证?”, 该问题的问题类型为查询某操作的适应证, 查询元路径: (a: 操作)-[:适应证]->(b: 疾病/临床所见)。

输入 cypher 语句:

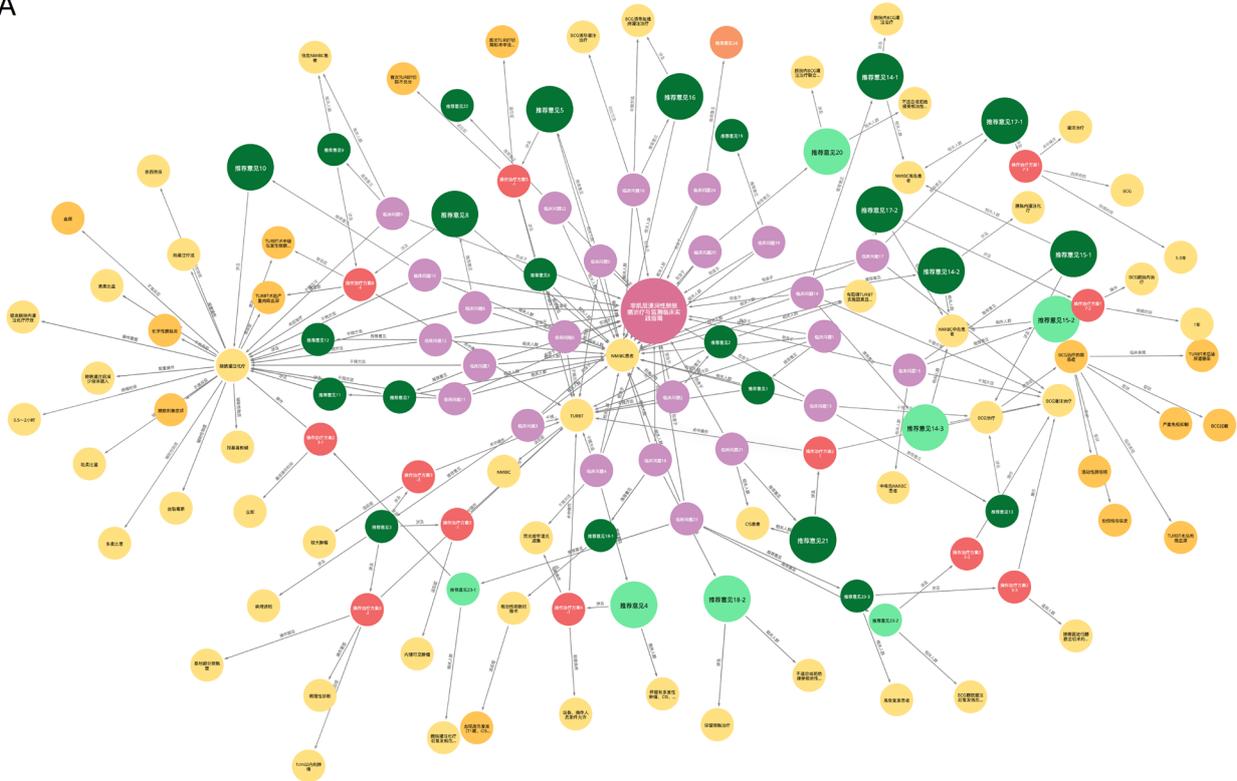
表5 知识抽取示例

Table 5. Examples of knowledge extraction

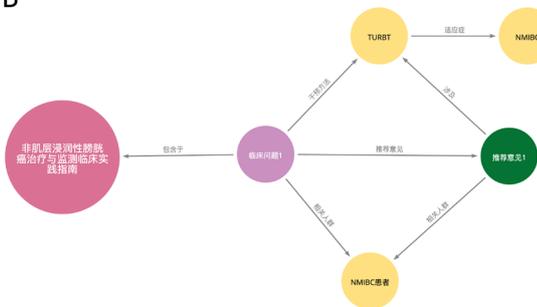
指南问题实体抽取	推荐意见实体抽取
【临床问题1】(临床问题): 【NMIBC患者】(人群) 行【TURBT】(操作)的适应证?	对【怀疑为NMIBC的患者】(人群), 【推荐】(文本) 【TURBT】(操作)作为诊断和初始治疗的方法
【临床问题2】(临床问题): 【NMIBC患者】(人群) 术后【即刻膀胱灌注化疗】(操作)的禁忌证?	【TURBT】(操作, 补充实体) 【术中发生膀胱穿孔】(临床所见)或【TURBT】(操作, 补充实体) 【术后严重肉眼血尿】(临床所见) 不宜行【即刻膀胱灌注化疗】(操作)
【临床问题3】(临床问题): 【低危患者】(人群) 术后是否只需行【即刻单次膀胱灌注化疗】(操作)?	对于【低危NMIBC患者】(人群), 术后可仅行【即刻单次膀胱灌注】(操作治疗方案)
【临床问题4】(临床问题): 对【NMIBC患者】(人群), 【BCG诱导加维持灌注治疗】(操作)是否优于【单纯BCG诱导灌注治疗】(操作)?	【NMIBC患者】(人群, 补充实体) 【推荐】(文本) 【BCG诱导加维持灌注治疗】(操作)

注: 【】中的文本为实体内容; ()中的文本为对应的实体所属概念

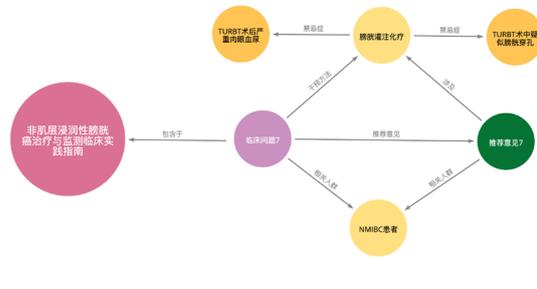
A



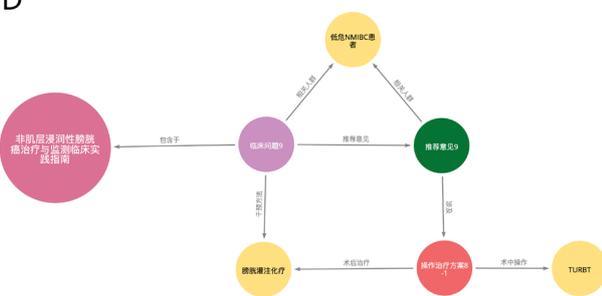
B



C



D



E

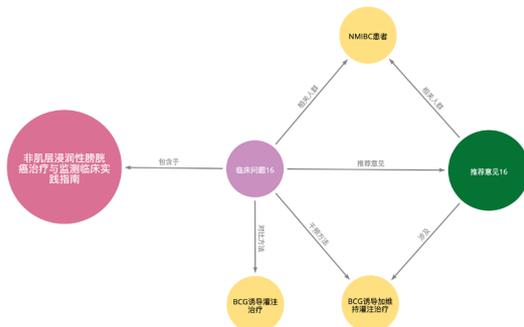


图4 NMIBC指南知识图谱

Figure 4. NMIBC guideline knowledge graph

注：A为NMIBC指南完整知识图谱；B、C、D、E分别为4个临床问题相连接的实体及相应的关系

match (a{name:'TURBT'})-[:适应证]->(b) where  
b: 临床所见 or b: 疾病 return b  
则可得到图 5A 的结果, 即 TURBT 的适应证  
为“NMIBC”。

## 2.2 查询某操作的禁忌证

对于临床问题 2: “术后即刻膀胱灌注化疗的禁忌证? ”, 该问题的问题类型为查询某操作的禁忌证, 查询元路径是: (a: 操作)-[:禁忌

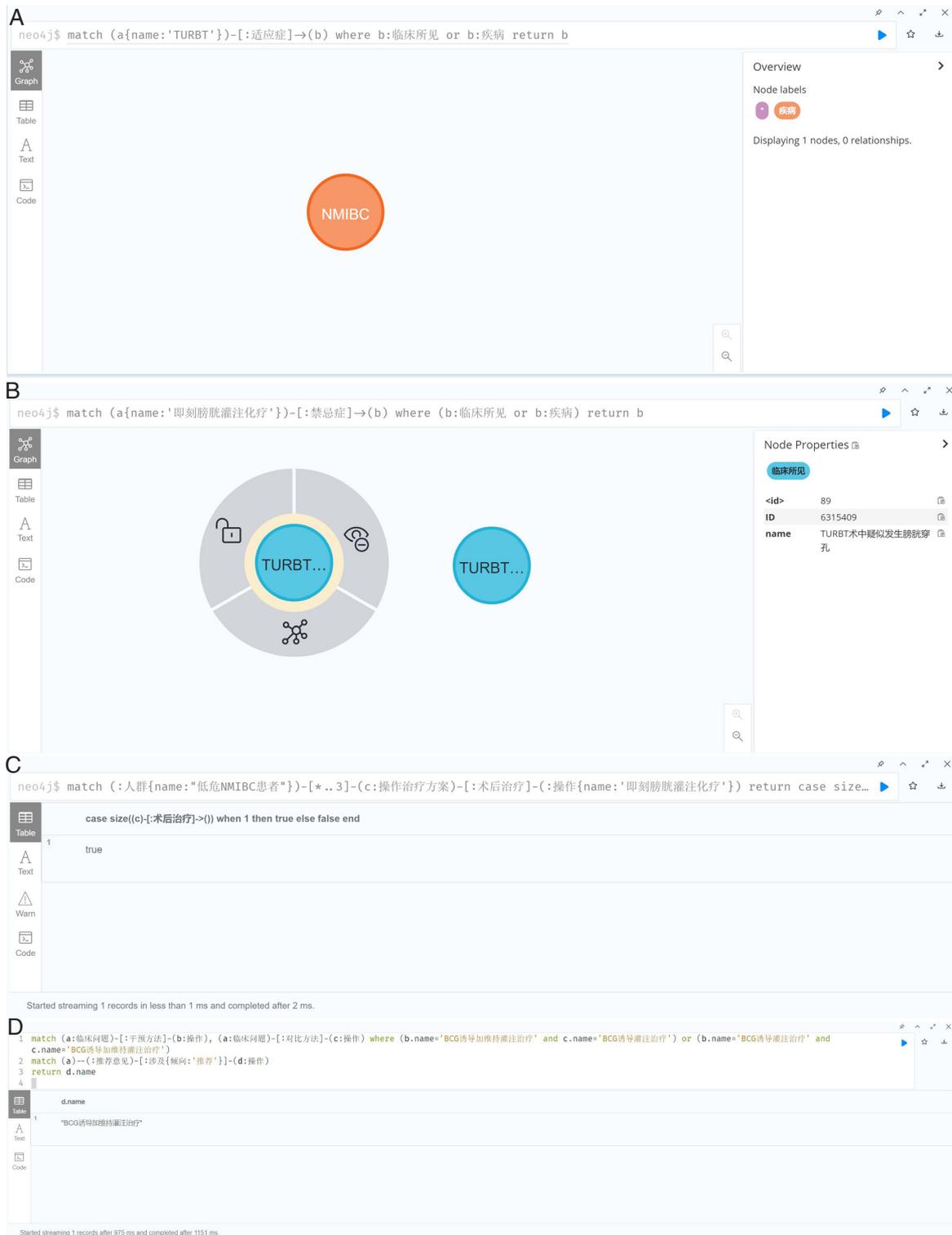


图5 NMIBC指南知识图谱临床实例验证

Figure 5. Validation of clinical examples of the NMIBC guideline knowledge graph

注: A为查询“临床问题1”的结果; B为查询“临床问题2”的结果; C为查询“临床问题3”的结果; D为查询“临床问题4”的结果

证 ]->(b: 疾病 / 临床所见)。

输入 cypher 语句:

```
match (a{name:'即刻膀胱灌注化疗'})-[:禁忌证]->(b) where (b:临床所见 or b:疾病) return b
```

则可获取图 5B 的结果, 即术后即刻膀胱灌注化疗的禁忌证为“术中发生膀胱穿孔”和“术后严重肉眼血尿”。

### 2.3 查询某人群的操作治疗方案中的术后操作及术后操作数量

对于临床问题 3: “低危患者术后是否只需行即刻单次膀胱灌注化疗?”, 该问题的问题类型为查询某人群的操作治疗方案中的术后操作及术后操作数量, 查询元路径: (a: 人群)-...-(操作治疗方案)-[:术后操作]->(b: 操作)。

输入 cypher 语句:

```
match (:人群{name:"低危 NMIBC 患者"})-[*..3]-[:操作治疗方案]-[:术后治疗]-[:操作{name:'即刻膀胱灌注化疗'})
```

return

```
case size((c)-[:术后治疗]->(o))
```

```
when 1 then true
```

```
else false
```

```
end
```

则可得到图 5C 的结果, 其中第一行代码检索术后治疗为即刻膀胱灌注化疗的操作治疗方案。随后检索操作治疗方案有几种术后治疗, 检索结果是只有一个, 所以回答是“true”。

### 2.4 查询指南推荐的操作方式

对于临床问题 4: “对 NMIBC 患者, BCG 诱导加维持灌注治疗是否优于单纯 BCG 诱导灌注治疗?”。该问题的问题类型为操作对比。

输入 cypher 语句:

```
match (a:临床问题)-[:干预方法]->(b:操作), (a:临床问题)-[:对比方法]->(c:操作) where (b.name='BCG 诱导加维持灌注治疗' and c.name='BCG 诱导灌注治疗') or (b.name='BCG 诱导灌注治疗' and c.name='BCG 诱导加维持灌注治疗')
```

```
match (a)-[:推荐意见]-[:涉及{倾向:'推荐'}]->(d:操作)
```

```
return d.name
```

则可得到图 5D 的结果, 其中第一行代码检索是否存在有将这两种操作方法互为对比的临床问题。随后寻找该临床问题对应的推荐意见。检

索结果是“BCG 诱导加维持灌注治疗”。

本研究后续会通过规则匹配的方法, 实现 NMIBC 系统的后台操作代码。当用户点击某种治疗方式或预后因素时, 系统会根据 NMIBC 知识图谱接收搜索条件内容, 通过规则匹配, 最后将其转换为 Cypher 语句并在 Neo4j 中查询答案。

本研究采用专家咨询法, 邀请专家对设计的 NMIBC 知识图谱进行评估以验证知识图谱的科学性与合理性, 并基于专家意见对知识图谱进行修改与补充, 确定最终版 NMIBC(单一)指南知识图谱。

## 3 讨论

本研究基于领域专家共识并参照 OMAHA Schema, 设计 NMIBC 领域的顶层本体, 提出并构建了 NMIBC 临床指南的知识图谱框架, 其二级概念架构及数据层(实体和关系)设计方法也可为其他疾病临床指南知识图谱的建设提供参考。本课题组考虑到过多的层数会使概念层庞杂而冗余, 增加运营与维护成本, 且过多的层数会使概念类别数量呈指数增加, 为实体抽取带来负担。另外, 由于实体提取算法的训练依赖于标注数据, 而少量标注数据无法覆盖过多概念类别。因此, 本课题组在实际操作中用二层概念层初步构建了该知识图谱, 发现其可以合理表达指南中的概念与语义关系。

本研究的知识图谱构建方法着重从知识源头对知识进行优化, 设计并调整概念间的关系结构, 对实体与实体间的关联进行细化完善, 并验证其可行性。对于指南中临床问题和推荐意见缺少的实体, 本研究进行了补充, 以使知识图谱节点更清晰, 查询或搜索时更方便快捷。另外, 在综合考虑指南数据存储以及为后续大量指南知识提取提供参考等情况后, 本课题组对临床问题根据 PICO 原则进行拆分, 然后由临床问题指向推荐意见, 从而确保指南信息存储的完整性。对指南的制订过程与方法、推荐意见中的操作说明和证据概述的补充, 以及对不同类型/版本指南的更新和融合将在本系列其他文章展示。

本研究仍存在一定局限。首先, 参照 OMAHA Schema 的语义结构规范虽然可以保证输入知识图谱的数据质量, 但仍需临床工作者和工程师对临床指南进行整理和总结, 对 Schema 中缺失的语

义类型、属性关系进行补充。其次，非结构化文本指南中有确定性知识和不确定性知识两种，不确定知识的选择和划分也是难点之一。更好的方法是引入多种知识表达方式以增强知识图谱的语义表达能力。

综上所述，本文提出的 NMIBC 指南的知识图谱框架设计不仅为指南数字化、智能化提供了基础，同样可以用于辅助临床决策，有利于指南的实施、推广与传播。本研究提出的框架设计不仅适用于 NMIBC 的知识图谱构建，对于其他医学专业或通用医学知识图谱的构建也可以提供思路。

### 参考文献

- Institute of Medicine (US) Committee on Standards for Developing Trustworthy Clinical Practice Guidelines. Clinical practice guidelines we can trust[M]. Washington (DC): National Academies Press, 2011. DOI: [10.17226/13058](https://doi.org/10.17226/13058).
- Heen AF, Vandvik PO, Brandt L, et al. A framework for practical issues was developed to inform shared decision-making tools and clinical guidelines[J]. J Clin Epidemiol, 2021, 129: 104–113. DOI: [10.1016/j.jclinepi.2020.10.002](https://doi.org/10.1016/j.jclinepi.2020.10.002).
- Vandvik PO, Brandt L, Alonso-Coello P, et al. Creating clinical practice guidelines we can trust, use, and share: a new era is imminent[J]. Chest, 2013, 144(2): 381–389. DOI: [10.1378/chest.13-0746](https://doi.org/10.1378/chest.13-0746).
- Siemieniuk RA, Agoritsas T, Macdonald H, et al. Introduction to BMJ rapid recommendations[J]. BMJ, 2016, 354: i5191. DOI: [10.1136/bmj.i5191](https://doi.org/10.1136/bmj.i5191).
- He J, Baxter SL, Xu J, et al. The practical implementation of artificial intelligence technologies in medicine[J]. Nat Med, 2019, 25(1): 30–36. DOI: [10.1038/s41591-018-0307-0](https://doi.org/10.1038/s41591-018-0307-0).
- Lu F, Cong P, Huang X. Utilizing textual information in knowledge graph embedding: a survey of methods and applications[J/OL]. IEEE Access 2020, 8: 92072–92088. DOI: [10.1109/ACCESS.2020.2995074](https://doi.org/10.1109/ACCESS.2020.2995074).
- Wang Q, Mao Z, Wang B, et al. Knowledge graph embedding: a survey of approaches and applications[J]. IEEE Transactions on Knowledge Data Engineering, 2017, 29(12): 2724–2743. DOI: [10.1109/TKDE.2017.2754499](https://doi.org/10.1109/TKDE.2017.2754499).
- Nicholson DN, Greene CS. Constructing knowledge graphs and their biomedical applications[J]. Comput Struct Biotechnol J, 2020, 18: 1414–1428. DOI: [10.1016/j.csbj.2020.05.017](https://doi.org/10.1016/j.csbj.2020.05.017).
- Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. CA Cancer J Clin, 2018, 68(6): 394–424. DOI: [10.3322/caac.21492](https://doi.org/10.3322/caac.21492).
- Richters A, Aben KKH, Kiemeny L. The global burden of urinary bladder cancer: an update[J]. World J Urol, 2020, 38(8): 1895–1904. DOI: [10.1007/s00345-019-02984-4](https://doi.org/10.1007/s00345-019-02984-4).
- 靳英辉, 曾宪涛. 中国非肌层浸润性膀胱癌治疗与监测循证临床实践指南(2018年标准版)[J]. 现代泌尿外科杂志, 2019, 24(7): 516–542. DOI: [10.3969/j.issn.1009-8291.2019.07.004](https://doi.org/10.3969/j.issn.1009-8291.2019.07.004). [Jin YH, Zeng XT. Evidence-based clinical practice guidelines for the treatment and monitoring of non-muscular invasive bladder cancer in China (2018 Standard Edition)[J]. Journal of Modern Urology, 2019, 24(7): 516–542.]
- OMAHA. OMAHA Schema[EB/OL]. (2021-11-12) [Access on 2021-12-08]. <http://wiki.omaha.org.cn/display/KG/OMAHA+Schema>.
- 张玲玉, 尹鸿峰. 基于 OAN 的知识图谱查询研究[J]. 软件, 2018, 39(1): 54–59. DOI: [10.3969/j.issn.1003-6970.2018.01.011](https://doi.org/10.3969/j.issn.1003-6970.2018.01.011). [Zhang LY, Yin HF. A Knowledge graph query algorithm based on OAN[J]. computer engineering & Software, 2018, 39(1): 54–59.]

收稿日期: 2021 年 10 月 23 日 修回日期: 2021 年 12 月 09 日  
 本文编辑: 桂裕亮 黄 笛

引用本文: 王永博, 高旷, 李绪辉, 等. 临床实践指南实施性促进研究之二: 基于非肌层浸润性膀胱癌指南的知识图谱框架设计 [J]. 医学新知, 2021, 31(6): 419–432. DOI: [10.12173/j.issn.1004-5511.2020111058](https://doi.org/10.12173/j.issn.1004-5511.2020111058)  
 Wang YB, Gao K, Li XH, et al. Research on promotion of implementation of clinical practice guidelines (II): framework design of knowledge graph construction based on guidelines for non-muscle invasive bladder cancer[J]. Yixue Xinzhi Zazhi, 2021, 31(6): 419–432. DOI: [10.12173/j.issn.1004-5511.2020111058](https://doi.org/10.12173/j.issn.1004-5511.2020111058)