



开放科学
(资源服务)
标识码
(OSID)

基于在线问诊记录的医疗疼痛感知知识图谱构建及查询应用

王世钰 崔超宁 由丽萍

山西大学经济与管理学院 太原 030006

摘要: [目的/意义] 基于医疗领域疼痛感知信息分析构建知识图谱, 为智能医疗服务优化提供参考。[方法/过程] 基于概念结构解析构建疼痛感知信息分类体系, 基于本体和语义学理论构建疼痛感知知识图谱; 应用 Fuseki 可视化工具和 SPARQL 查询语言, 将知识图谱应用于推理式语义查询。[结果/结论] 通过在线问诊记录验证, 证明了知识图谱查询的有效性和全面性, 是知识图谱在医疗健康领域的一次有益探索。

关键词: 知识图谱; 疼痛感知信息; 本体

中图分类号: G35

Construction and Query Application of Medical Pain Perception Knowledge Graph Based on Online Reviews

WANG Shiyu CUI Chaoning YOU Liping

College of Economics and Management, Shanxi University, Taiyuan 030006, China

Abstract: [Objective/Significance] Based on the analysis of pain perception information in the medical field, a knowledge graph is constructed to provide a reference for the optimization of intelligent medical services. [Methods/Processes] In this study, we construct a pain perception information classification system model based on conceptual structure analysis, and a pain perception knowledge graph based on ontology and semantic theory, using visualization tool Fuseki and query language SPARQL to apply the knowledge graph to reasoning semantics inquire. [Results/Conclusions] Verification of online medical Consultation records proves the effectiveness and comprehensiveness of knowledge map query, which is a useful exploration of knowledge map in the field of medical and health.

Keywords: Knowledge graph; pain perception information; ontology

基金项目 国家自然科学基金面上项目“面向汉语篇章语义分析的框架推理技术研究”(61772324)。

作者简介 王世钰(1994-), 硕士, 研究方向为网络数据挖掘; 崔超宁(1995-), 硕士, 研究方向为网络数据挖掘; 由丽萍(1979-) 博士, 教授, 硕士生导师, 研究方向为网络数据挖掘, E-mail: yoliping@163.com。

引用格式 王世钰, 崔超宁, 由丽萍. 基于在线问诊记录的医疗疼痛感知知识图谱构建及查询应用[J]. 情报工程, 2022, 8(5): 65-73.

引言

随着信息技术在医疗卫生领域的广泛应用,众多医疗网站及 APP 开始涌现,且积累了大量疼痛相关的问诊记录信息。由于该类信息对于患者疼痛感知的表达更为准确与全面,因此获取及分析此类信息可以为高效判断患者的病情及提高医疗服务质量提供决策支持。

知识图谱(Knowledge Graph)是一种基于关系与结构化的知识表达方式。既可以与不同的实体相连接,实现“实体-关系-实体”的三元组,以表示实体之间的关系,也可以连接实体与其相关属性-值对,从而构成网状的知识结构;并根据用户的需求,运用可视化技术以图片形式向用户反馈结构化的知识,更为直观的展示出其各种关系,使用户不必自行从大量信息中进行梳理及筛选,就能获取准确的信息。本体(ontology)^[1]提供了特定领域中概念或对象类型及其属性和相互关系的表示,可实现语义网中的信息及关系的结构化描述,是知识图谱的一种信息描述形式,并且有助于计算机处理效率的提升。因此,知识图谱可以将信息存储与语义相融合,有效的组织疼痛感知信息,从而完善语义推理机制对于医疗领域知识的应用。

国外对于知识图谱在医疗领域的应用与研究均早于国内。当前具有最大影响力的医疗知识库 UMLS^[2]即通过知识图谱的方式对医学各概念之间的关系进行描述,其中包含了 100 多万个医学概念,133 类医学领域语义类型、54 种语义关系。在 UMLS 的基础上,中国医学科学院医学信息研究所开发研制了中文一体化医学语言系统 CUMLS^[3]。国外的研究工作主要

集中在对于医疗领域实体知识及关系知识的抽取^[4-6],而国内相关研究则以特定范围或特定载体信息为主^[7-12]。目前,总体上我国缺乏对某一具体领域语义层面的深入研究,就医疗领域知识图谱的相关研究仍处于起步阶段。

在疼痛感知信息分析领域中,由 Melzack^[13]提出的 McGill 疼痛问卷(McGill Pain Questionnaire, MPQ)是目前常用的临床及研究工具。但由于制定年代久远而有一定的局限性,Dworkin 等^[14]基于简版 MPQ(Short-form MPQ, SF-MPQ)制定了 SF-MPQ-2。随后中国各地 10 家三甲医院的疼痛科在中华医学会疼痛学分会倡导下相互合作,制定与验证了中文版 SF-MPQ-2^[15]。SF-MPQ 和 SF-MPQ-2 为疼痛感知信息分析提供了专业的概念框架和参考标准,但其中的信息太过概括。由于疼痛感知作为一种在患者体验的基础上所得到的柔性信息,具有较强的主观性,因此,用这种量表描述患者的疼痛信息在描述深度上是远远不够的。随着在线医疗应用的蓬勃发展,网络上积累了大量医患之间的对话信息,其具有真实性、全面性和语义丰富性,可以作为疼痛感知信息描述的重要知识源,但这一点在学术界还没有引起足够的重视。同时,随着自然语言处理和语义网相关技术日趋成熟,目前基于问诊记录的知识获取和信息查询具有高度的可行性。

因此,本文结合在线医疗问诊记录相关验证数据与医疗领域疼痛感知专业知识,结构分析患者疼痛信息;基于自然语言处理和语义网相关技术分析在线医疗问诊记录,实现疼痛感知的知识获取;运用知识图谱工具描述和组织疼痛感知信息,实现相关知识的表示与存储;

最后，在应用方面通过语义推理查询疼痛感知知识图谱。

1 疼痛感知信息分析及知识图谱构建

1.1 疼痛感知信息分析

1.1.1 疼痛感知定义及类型

不同种类的疼痛是人主观感觉的集成，对疼痛感知信息的全面描述，不仅要科学分类疼痛感知的性质，还要描述疼痛感知的部位、程度等特征。本文从文献 [15] 中获取了中文版 SF-

MPQ-2 的疼痛感知的相关分类,结果如图 1 所示。

1.1.2 疼痛感知属性

本文以医疗领域公认综合疼痛评估的关键组成部分的标准^[16]为主，同时咨询相关领域医学专家，共同作为疼痛感知相关属性定义的主要参考。由于疼痛感知信息主要以语言作为交际媒介在医患之间进行相互传递，因此，语言中蕴含了相应的属性信息。所以，参考语言学领域关于感知场景的相关论述^[17]，将身体感知的语义角色类型——身体部位、程度、原因、时间等也纳入描述范围。疼痛感知属性定义，如表 1 所示。

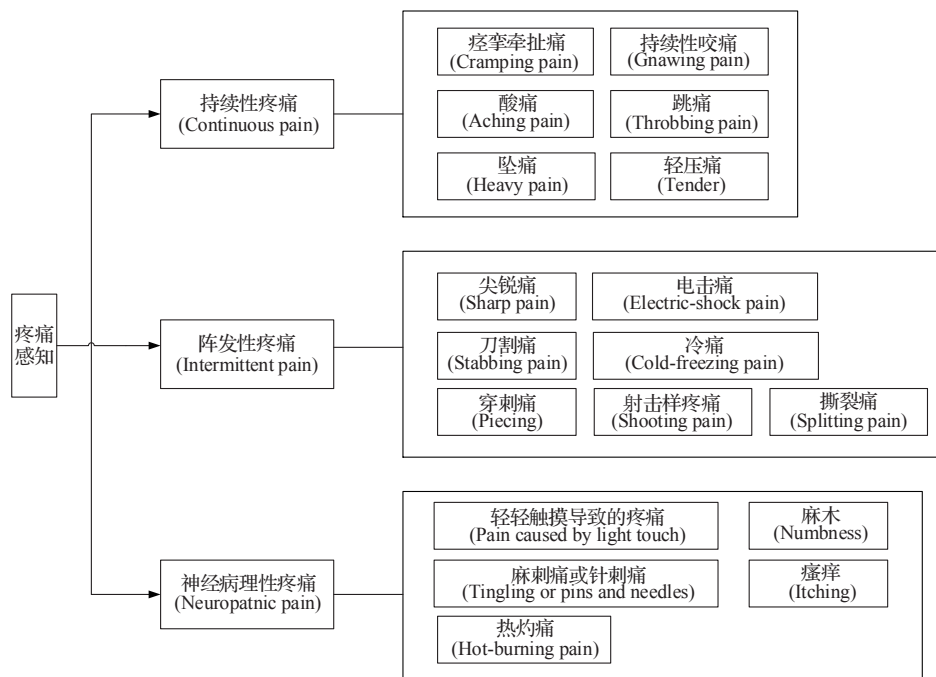


表 1 疼痛感知属性定义表

疼痛感知属性	定义
疼痛部位	产生疼痛的部位。
程度	描述疼痛的强度，根据综合疼痛评估标准以十级为描述刻度。
质量	描述疼痛的性质，即疼痛类型。
持续性	描述疼痛在时间上的持续情况。
日变化	描述疼痛程度于一日内的强弱变化趋势。

1.1.3 疼痛感知因果关系

造成疼痛程度变化的因素分为缓解因素和加重因素,其对于医生评估病情十分重要。同时,疼痛的产生必然伴随着对自身生活及情绪上的影响。因此,疼痛感知变化原因及其影响对于疼痛感知信息分析是必不可少的部分,其定义如表2所示。

表2 疼痛感知因果关系定义表

疼痛感知因果关系	定义
缓解因素	疼痛程度减轻的相关因素,主要体现为行为或活动。
加重因素	疼痛程度加重的相关因素。
对情绪的影响	疼痛给对患者情感及心理状态的影响。
对生活的影响	疼痛对患者身体功能及生活方式的影响。

1.1.4 检查和诊断信息

问诊过程中的疼痛感知信息,即可以体现出描述疼痛感知状态的细化信息,还可以体现出患者的病史信息、医生的检查建议及诊断结果,其定义如表3所示。

表3 检查和诊断信息定义表

检查和诊断信息	定义
检查建议	医生对于患者此次疼痛状况所提出的相关医疗检查项目。
诊断结果	医生就症状检查结果及患者描述,判断病人病情及疾病类型。
病史	患者于此次疼痛之前患过的疾病,包括疾病类型和治疗方法。

1.2 疼痛感知知识图谱构建

1.2.1 疼痛感知概念模型设计

在概念模型设计上,首先确定本体顶层大

类,即具有共同属性和行为的对象集合,包括患者、疼痛部位、疼痛类型、疼痛程度和病史,且将其设定为同级关系(Sibling Class);在此基础上,采用自顶向下的方法,参考医患问诊对话记录,对上层概念逐步细化,为顶层类添加子类(Subclass),从而建立类与类之间及类与自身属性之间的关系,即定义对象属性和数据属性,数据属性包括:患者年龄、性别、疼痛影响因素等。结果如图2所示。

1.2.2 疼痛感知知识获取

(1) 语义分类词表构建

由于患者缺乏医学专业知识,所以,多采用自然语言表达方式描述病情,而非医疗专业术语。因此,需构建能够与医疗领域疼痛感知类型进行匹配的专业词表。本文以在线问诊量比较大且疼痛现象比较多的颈椎病为例,从好大夫在线(<https://www.haodf.com>)爬取医患对话记录2500条,采用哈工大语言技术平台(LTP)进行文本预处理,再以形容词和动词为主进行人工筛查,初步实现疼痛类型词语的提取;继而利用语言学词典《新编同义词词林》和《现代汉语分类词典》等补充相关同义词以便丰富词表。共收集了89个词语,按疼痛类型进行语义分类,结果如表4所示。

(2) 疼痛等级词表构建

目前,对疼痛程度的描述医疗领域普遍采用的是十级量表,即0代表没有疼痛,10代表疼痛最严重,1到10代表疼痛程度依次加重。但在问诊记录中患者描述症状时通常使用的是程度副词,如“一点点”“有点”“不严重”“很”“厉害”等。由于语言中的程度词与十级量表并不能够相互对应,为分析文本中的疼痛程度等级

信息，则需通过语义分析识别程度词，并根据十级量表对其进行合理赋值。因此，本文将十级量表调整为三个区间以对应程度词，如表 5 所示。

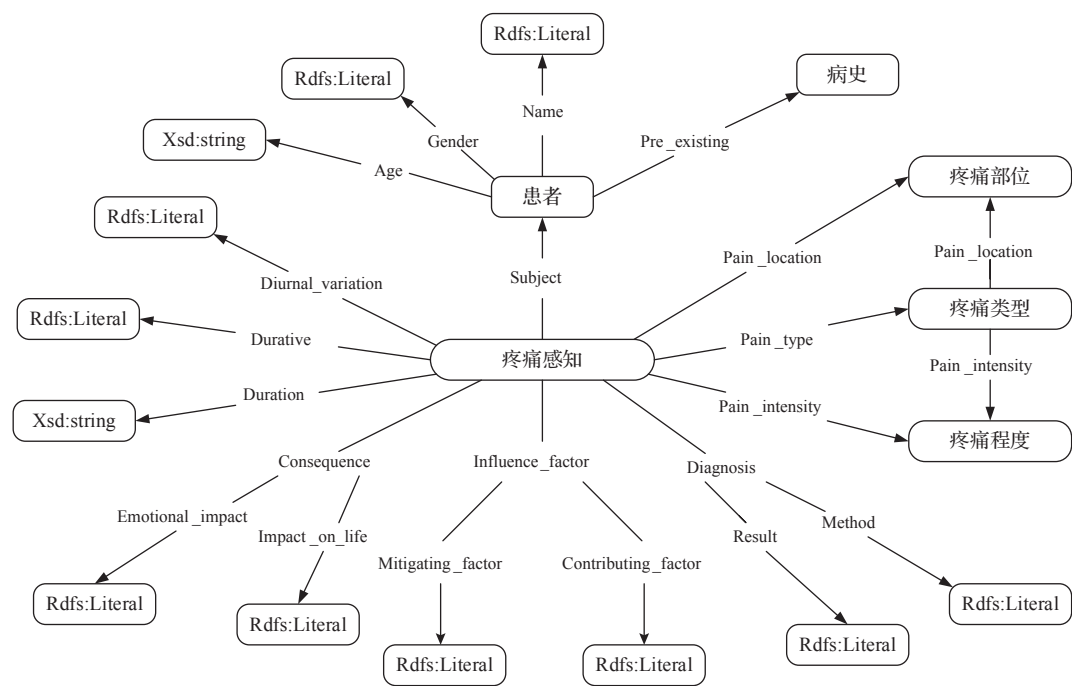


图 2 本体类与属性结构图

表 4 疼痛感知词表

上层类	下层类	词汇
持续性疼痛	轻压痛(Tender)	压痛
	钝痛(Dull pain)	疼、痛、疼痛、神经痛、肌肉疼、皮肤疼、偏头疼
	胀痛(Distending pain)	胀、胀疼、闷胀、胀痛、发胀、肿胀、鼓包、鼓胀、红肿、虚肿、肿大、水肿、浮肿
	痉挛牵扯痛(Cramping pain)	抽、抽搐、抽筋、痉挛、拘挛、抽溺、转筋、掣痛、放射性疼痛、反射痛、放射痛、拉扯疼
	酸痛(Aching pain)	酸、酸疼、酸痛、酸胀、酸困、发酸、酸软、泛酸、酸沉感
	跳痛(Throbbing pain)	窜痛、跳痛、跳动、一跳一跳
阵发性疼痛	刀割痛(Stabbing pain)	绞痛
神经病理性疼痛	麻木(Numbness)	僵硬、发僵、僵直、麻、发麻、发木、麻木、酥麻、胀麻、酸麻、发僵、麻痹、麻酥酥、麻痛、麻疼、放射性麻木、晕、束带感、昏、电麻感、眩晕、发沉、发闷、发紧
	瘙痒(Itching)	痒、发痒、痒痒、刺痒、刺挠、奇痒
	热灼痛(Hot-burning pain)	烧、灼痛、灼热、灼伤、烫、烫伤
	轻轻触摸导致的疼痛(Pain caused by light touch)	触痛、触碰疼痛
	麻刺痛/针刺痛(Tingling or pins and needles)	扎痛、刺痛、刺疼、针刺感

表 5 疼痛程度词与十级量表对应表

疼痛十级量表	疼痛程度词
7-10	绝、格外、超、超级、贼、忒、非常、爆、痛不欲生、至极、要命
4-6	很、非常、明显、特、特别、极、极其、十分、强烈
1-3	一点、有点、略微、隐隐、挺、蛮、相当、比较、稍微、不严重、隐、约约、轻度

(3) 本体实例添加

由于类和属性与语言表达中的语义角色相对应,因此,需对文本进行语义角色标

注。本文利用课题组自主研发的语义标注软件,即“基于框架本体的中文情感语义标注与查询系统”(软件著作权登记号:2018SR823004),标注出文本中的疼痛部位、程度、因果关系、时间等角色。以此为基础,将所标注出的疼痛感知知识作为实例,添加至疼痛感知本体中。其中,主要类包括疼痛部位、疼痛类型、病史等,主要属性包括年龄、性别、疼痛时长等,共 97 个实例,其实例添加结果如表 6、表 7 所示。

表 6 主要类的实例添加结果表

类	实例
疼痛部位	肩周、背、颈椎、太阳穴、肩、胳膊、脖子、四肢、手指、眼、头、心脏、心前区、臂、脚
病史(患病类型)	贫血、三高、肩周炎、颈椎病、高血压、类风湿关节炎、糖尿病、冠心病
病史(治疗方式)	常年吃药、颈椎病手术、推拿疗法、心脏支架/搭桥手术

表 7 主要属性的实例添加结果表

属性	实例
持续性	一直、断断续续、偶有、偶尔、持续
日变化	一直一样疼、一天下来加重、早/中/晚最疼、早/中/晚轻点、早晨起床时好些
疼痛时长	最近、半年、一年、两年、大于半年、一个月、两个月
缓解因素	按摩后、躺着、按住、转移注意力(看电视、听音乐等)
加重因素	阴天时、轻压、站久了、运动后、坐着、躺多了、触碰、洗衣服时
对情绪的影响	紧张、害怕、忧惧、忐忑、不安、发毛、恐惧、惶恐、焦躁、深恐、厌烦、厌恨、令人厌恶的、抱怨、埋怨、折磨、被惩罚、煎熬、恶心、嫌恶
对生活的影响	四肢乏力、不能干活、睡不着、眼花、活动受影响、坐轮椅、半夜疼醒、一夜不能睡、走路有一点点无力
检查方式	核磁共振、CT、超声、MRI、吃药、X 线检查

1.2.3 疼痛感知知识存储

疼痛感知知识图谱存储格式为 RDF 格式。RDF 标准语法基于可扩展标记语言(Extensible Markup Language, XML)进行设计,利用 XML 文件的开头命名空间对不同领域资源进行划分,同一领域实体和关系 URI 前缀相同,将前缀定义为 URIfref,即标签术语。利用标签术

语与 URI 中的关键词对资源进行描述,语句简单清晰,有助于提高计算机的理解效率。例如以下问诊记录:

患者年龄: 48 岁
患者性别: 男
病情描述: 两个手指指尖有点发麻,运动后发麻程度加重,一个多月了,无其他症状。有高血压,常年吃抗压药。
希望提供的帮助: 请问需要做什么检查? 问题严重吗?
医生回复: 可以去做核磁共振或四肢彩超,检查结果出来才能判断严不严重。

用 RDF/XML 语言对上述例子中疼痛感知本体进行描述, 如图 3 所示:

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:contact="http://www.semanticweb.org/administrator/ontologies/2019/2/untitled-ontology-40#">
  <contact:疼痛感知文本 1 rdf:about="http://www.semanticweb.org/administrator/ontologies/2019/2/untitled-ontology-40#疼痛感知文本 1">
    <contact:pain_location>手指</contact:pain_location>
    <contact:pain_type>麻</contact:pain_type>
    <contact:pre_existing>高血压</contact:pre_existing>
    <contact:pre_existing>吃药</contact:pre_existing>
    <contact:method rdf:resource="核磁共振或四肢彩超"/>
    <contact:durative rdf:resource="一个月"/>
    <contact:contributing_factor rdf:resource="运动后"/>
  </contact:疼痛感知文本 1>
</rdf:RDF>
```

图 3 RDF/XML 语言描述疼痛感知本体

其中, 用“contact”和标签术语 (pain_location 等) 描述数据的具体内容。例如: “contact”和“疼痛感知文本 1”以关键词“about”确认其 URI, contactant: pain_location 表示在该 URL 下的疼痛部位是“手指”。由于在一个 RDF 文件中可能存在多个“手指”, 利用 URI 即可表示特定的“手指”。

将问诊记录中的实例相对应于其不同的类和属性, 实现疼痛感知的知识存储。上述例子

中知识存储如表 8 所示。

表 8 问诊记录实例对应的类与属性表

类	实例	属性	实例
疼痛部位	手指	年龄	48
疼痛类型	麻	性别	男
疼痛程度	1—3	疼痛时长	一个 (多) 月
病史 (患病类型)	高血压	加重因素	运动后
病史 (治疗方式)	吃抗压药	检查方式	核磁共振或四肢彩超

1.2.4 疼痛感知知识图谱可视化

利用 Protégé5.2.0 进行可视化, 其插件 OntoGraf 支持中文本体多种形式的展示, 利于理解整体结构及类之间、实例之间的语义关联, 将知识直观地展示给用户, 满足读取和查询的需求。本文构建的疼痛感知领域知识图谱包括 5 个类, 10 个属性, 13 种实体间关系以及 182 个实例, 部分展示如图 4。

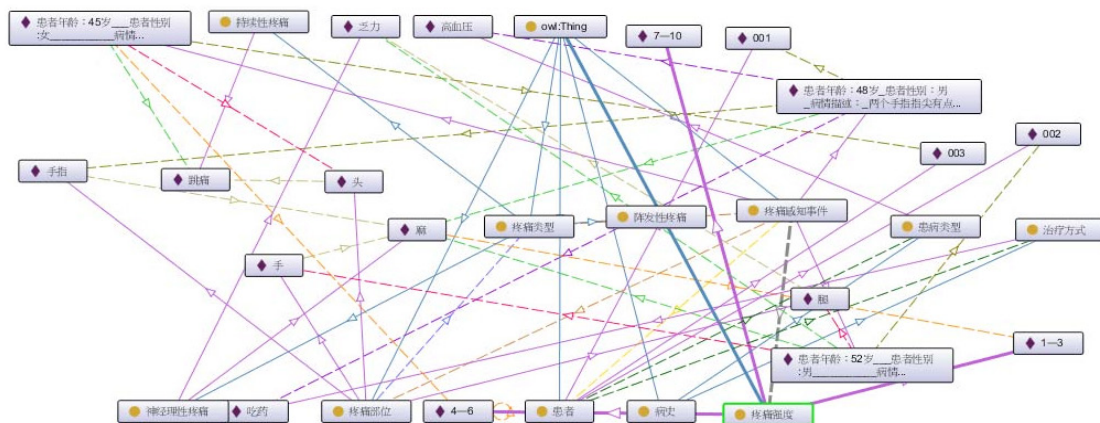


图 4 疼痛感知本体中部分类与属性关系图

2 疼痛感知知识图谱查询应用

本文应用可视化工具 Fuseki 以及查询语言 SPARQL (Simple Protocol and RDF Query Language), 对疼痛感知知识图谱进行推理式查

询。SPARQL 结合了 rdfDB、RDQL 和 SeRQL 等 RDF 查询语言的优势, 可以在语义层面进行检索, 进而满足用户多种查询需求, 与传统的 SQL 语句相比, 提高了数据查询的有效性和全面性。

知识图谱按照三元组的方式对信息进行存储,为各实例之间加入多种语义关系,提供了不同的关系链种类,利于搜索结果的发散性。传统的关键词检索结果多而杂,用户需对检索结果自行甄别和筛选,从而可能导致遗漏信息。然而,知识图谱提升了信息的密度及关联度,可降低检索结果的单一性,并向用户反馈结构化的知识,使用户可以准确定位和深度获取知识,进一步提高检索效率,但由于医学领域的专业性,在其应用方面仍需基于专业人员的判断进一步加以考量。

3 结束语

本文通过对医疗领域疼痛感知信息分析从而构建知识图谱,并进行查询应用,具体研究贡献包括:(1)结构分析患者疼痛感知信息,并基于自然语言处理技术构建了医疗领域疼痛感知词表。(2)利用知识图谱工具及相关技术构建了疼痛感知知识图谱。(3)将疼痛感知知识图谱应用于语义查询,实验结果表明基于知识图谱的检索具有较高的效率和准确性,具有一定的应用价值。

本文研究结果提供了计算机可读、可理解的疼痛感知知识资源,从技术层面构建了疼痛感知专业领域知识图谱,并于一定程度上实现了推理式语义查询,为医疗健康领域知识图谱的应用提供借鉴,但仍存在一定的局限性,即在该知识图谱的专业应用方面仍需基于医学专业人员的判断进一步加以考量。

参考文献

- [1] 邓志鸿,唐世渭,张铭,等. Ontology 研究综述[J]. 北京大学学报(自然科学版), 2002(5):730-738.
- [2] Olivier B. The Unified Medical Language System (UMLS): integrating biomedical terminology[J]. Nucleic Acids Research, 2004(1):267-270.
- [3] 李丹亚,胡铁军,李军莲,等. 中文一体化医学语言系统的构建与应用[J]. 情报杂志, 2011, 30(2):147-151.
- [4] Li H D, Tang B Z, Chen Q C, et al. HITSZ_CDR: an end-to-end chemical and disease relation extraction system for BioCreative V[J]. Database, 2016: baw077.
- [5] Liu S Y, Tang B Z, Chen Q C, et al. Drug-Drug Interaction Extraction via Convolutional Neural Networks[J]. Computational and Mathematical Methods in Medicine, 2016:1-8.
- [6] Kavuluru R, Rios A, Tran T. Extracting Drug-Drug Interactions with Word and Character-Level Recurrent Neural Networks[C]. IEEE International Conference on Healthcare Informatics. IEEE, 2017:5-12.
- [7] 李国栋. 基于文献定量及知识图谱方法的医学信息学研究现状分析[J]. 中国科技信息, 2013(24): 125-127.
- [8] 化明艳,高月,李玉坤.《医疗卫生装备》杂志知识图谱分析[J]. 医疗卫生装备, 2013, 34(1):91-94.
- [9] 刘俊丽,张秀梅,蒋勇青. 基于文本挖掘的乙型肝炎相关文献知识图谱分析[J]. 医学信息学杂志, 2014, 35(1):48-53.
- [10] 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+59.
- [11] 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.
- [12] 龙宇晓,李筑艳. 当代中国苗族医药研究文献的知识图谱分析[J]. 贵州大学学报(社会科学版), 2016, 34(3):81-90.
- [13] Melzack R. The McGill Pain Questionnaire: Major properties and scoring methods[J]. PAIN, 1975, 1(3): 277-299.
- [14] Dworkin R H, Turk D C, Revicki D A, et al. Development and initial validation of an expanded and revised version of the Short-form McGill Pain Questionnaire (SF-MPQ-2)[J]. Pain, 2009, 144(1-2): 35-42.
- [15] 李君,冯艺,韩济生,等. 中文版简版 McGill 疼痛问卷-2 的制定与多中心验证[J]. 中国疼痛医学杂志, 2013, 19(1):42-46.
- [16] Murray P, Barker D. C4 An Institutional Comprehensive Pain Assessment: WILDA Revised[J]. Pain Management Nursing, 2018, 19(2):106.
- [17] Fillmore C J. Frame semantics and the nature of language[J]. Annals of the New York Academy of Sciences, 1976, 280(1):20-32.