

国内外生物医学领域本体研究与 实践进展*

赵洁¹ 司莉²

(1. 山西财经大学信息学院, 太原 030006; 2. 武汉大学信息资源研究中心, 武汉 430072)

摘要: 本文通过文献调查法归纳总结近十年(2010—2019年)国内外生物医学领域本体文献, 简要介绍国内外重要生物医学领域本体, 从构建、整合、扩展3个方面归纳总结生物医学领域本体的相关研究, 以及其在医学信息描述、查询扩展、知识库构建、术语相似度计算等方面的相关实践, 并对未来发展趋势进行展望, 以期为我国生物医学领域本体的构建与应用提供借鉴。

关键词: 生物医学领域; 本体; 本体构建; 本体整合; 本体扩展; 本体应用

中图分类号: G254 **DOI:** 10.3772/j.issn.1673-2286.2020.08.002

引用格式: 赵洁, 司莉. 国内外生物医学领域本体研究与实践进展 [J]. 数字图书馆论坛, 2020 (8) : 7-14.

本体是对共享概念体系的形式化规范说明, 具有语义准确、概念清晰、语义关系丰富、成果开放集成等优势, 被广泛应用于信息检索、知识管理等领域, 用以促进、支持和提高资源发现和信息检索的质量^[1]。生物医学领域是涉及从生物体到化学结构等多个学科的复杂领域, 对其知识结构及关联信息资源的组织是实现相应资源有效获取的重要条件。而本体作为包含概念与概念关系的语义化知识组织工具, 具有良好的结构化表示信息的能力和与人机可读的特点, 成为生物医学领域资源组织的主要方式。

生物医学领域本体是以解决生物医学术语的通用性和规范性, 以及构建医学健康术语间关系为目的的领域本体, 其作用是促进信息系统理解生物医学领域的语言含义。有些学者对生物医学领域本体的研究成果进行了梳理。如陈焱等^[2]介绍了统一医学语言系统(Unified Medical Language System, UMLS)、医学术语系统命名法——临床术语(Systematized Nomenclature of Medicine-Clinical Terms, SNOMED CT)、人类发育解剖学本体(Ontology of Human Developmental Anatomy, HUMAT)、基因本体(Gene

Ontology, GO)、医学知识库(NKIMed)、中文一体化医学语言系统(Chinese Unified Medical Language System, CUMLS)等国内外医学领域的主要本体。吴明智等^[3]通过文献调查, 分析了当前生物医学本体的研究现状与研究热点。张庆等^[4]通过相关文献高频主题词的共现聚类分析, 总结生物医学本体在生物信息学、临床医学、医学信息学以及人工智能中的应用情况。吴正荆等^[5]从开发时间、覆盖内容、结构等方面对UMLS、开放生物医学本体(Open Biomedical Ontology, OBO)、临床医学本体OpenGALEN(Open Generalized Architecture for Languages, Encyclopedias and Nomenclatures in Medicine)、生物医学调查本体(Ontology for Biomedical Investigations, OBI)4类生物医学本体进行深入比较, 总结目前本体开发中存在的同义词与术语表达、对应关系与间隔尺度等主要问题。

近年来, 生物医学领域本体的研究与实践取得了新的进展, 亟需全面梳理与系统综述, 以应对当前大数据环境下医学知识表示与组织的需求。本文首先对现有国内外常用生物医学领域本体进行归纳总结; 其次, 通过对2010—2019年国内外相关文献调研, 梳理近十

*本研究得到山西财经大学青年科研基金项目“基于社会化问答的用户健康知识结构研究”(编号: QN-202017)资助。

年国内外生物医学领域本体的研究与实践；最后，展望生物医学领域本体的未来发展。通过上述3个方面的研究，期望能够为我国生物医学领域本体的构建与应用提供借鉴。

1 国内外主要生物医学领域本体概览

生物医学领域本体发展至今，已形成较为成熟、不同规模、不同生物医学子领域的本体，被广泛应用于创建、处理、检索和整合生物医学数据和信息。笔者对国

内外比较常用的生物医学领域本体进行归纳总结，以促进生物医学领域知识的获取与利用。

1.1 国外主要生物医学领域本体概览

国外生物医学领域本体包括UMLS语义网络、SNOMED CT、GO、疾病本体（Disease Ontology, DO）、人类表型本体（Human Phenotype Ontology, HPO）、临床药品标准命名术语表（RxNorm）等。表1是国外代表性生物医学领域本体概览。

表1 国外代表性生物医学领域本体概览

本体名称	本体概述	创建单位	本体类型	本体规模	本体结构
UMLS 语义网络	顶层本体，为所有的UMLS概念提供整体概念框架	美国国立医学图书馆 (National Library of Medicine)	开放医学领域本体	135种语义类型，54种语义关系	语义类型 (Semantic Type)、语义关系 (Semantic Relations)
SNOMED CT	世界上最全面、准确的多语言健康术语系统，协助临床健康信息的电子交换，可以映射到ICD-9和ICD-10等编码系统，有助于促进语义互操作，被超过50个国家接受为全球通用的健康术语	美国病理家学会 (College of American Pathologist)	开放医学领域本体	30多万条医学概念，700多万条语义关系	概念表 (Concepts Table)、术语表 (Terms Table)、层次和关系表 (Hierarchies & Relationship Table)、描述逻辑 (Description Logic)
GO	提供当前从人类到细菌的许多不同生物体基因功能的科学知识的计算表示	基因本体联盟 (Gene Ontology Consortium)	基因本体	50 217个术语	-
DO	对人类疾病的名称、表型特征进行本体化描述，被用于统一模型生物的疾病注释，并改进生物学与临床人类疾病相关数据的互操作性，同时也被用于识别与疾病相关的蛋白质	美国西北大学基因药物中心、马里兰大学医学院基因组科学研究所	疾病本体	-	生物学过程本体 (Biological Process)、分子功能本体 (Molecular Function)、细胞成分本体 (Cellular Component)
HPO	为遗传性、先天性和获得性疾病的表型特征提供结构化和受控的词汇表	美国国立卫生研究院 (National Institutes of Health)	领域本体	超过1.3万个术语、超过15.6万个遗传疾病注释	表型异常本体 (Phenotypic abnormality)、遗传模式本体 (Mode of Inheritance)、临床修饰符本体 (Clinical modifier)、临床过程本体 (Clinical course)、频率本体 (Frequency)
RxNorm	为药品提供规范化的名称和唯一的标识符，使计算机系统能够有效和明确地交流药物相关信息	美国国立医学图书馆	药物本体	-	-

1.2 国内主要生物医学领域本体概览

相较于国外生物医学领域本体,国内发展较慢。

国内生物医学领域本体包括CUMLS、中医药学语言系统(Traditional Chinese Medicine Language System, TCMLS)等。表2是国内代表性生物医学领域本体概览。

表2 国内代表性生物医学领域本体概览

本体名称	本体概述	创建单位	本体类型	本体规模	本体结构
CUMLS	用以加强中文生物医学资源的自动化、规范化处理能力,为中文生物医学资源中语言异构性、中文生物医学资源在系统中的分散性、跨系统资源描述的不一致性等对生物医学信息检索带来障碍的问题提供解决方案。应用于医学文本自动处理、医学文献信息检索系统	中国医学科学院医学信息研究所	通用本体	包含10余个生物医学领域资源	医学词表、语义网络、构建工具
TCMLS	中医药学及其相关学科的综合语言系统,规范定义相关概念术语,并用于提高计算机理解相关词汇和语义的能力,可用于数据处理、知识检索、自然语言处理、专业词表编制、本体建设等	中国中医科学院中医药信息研究所	领域本体	10万个概念、30万个术语、127万条语义关系	基础词库、语义网络

2 生物医学领域本体的相关研究

本文采用文献调研法获取与本研究主题相关的近十年研究成果,即2010—2019年。具体的检索策略如下。

①国外文献检索策略。检索数据库为Web of Science,检索式为TS=(("medical" OR "health" OR "disease") AND "ontology") AND SU=Information Science & Library Science,检索时间截至2020年3月15日,检索结果有84篇,经主题筛选与全文获取要求,其中25篇为本文研究对象。②国内文献检索策略。检索数据库为中国知网,检索主题词包括疾病本体、医学本体、药物本体,检索时间截至2020年3月20日,检索结果有129篇,经主题筛选与全文获取要求,其中54篇为本文研究对象。笔者以这79篇文献为研究对象进行文献综述,梳理近十年国内外生物医学领域本体的应用情况、研究热点与未来发展趋势。

基于对上述文献的归纳总结,生物医学领域本体的相关研究集中在构建、整合与扩展3个主题。笔者分别对这3个主题的相关研究进行述评。

2.1 生物医学领域本体的构建

除应用现有成熟的生物医学领域本体之外,不同学者针对应用需要,建立了肿瘤、神经系统、肝脏疾病、产科和新生儿、孕产、伤口管理、军事医学等特定医

学领域的本体,并通过融合其他来源的知识,使本体更能体现公众关心的关键问题。针对本体构建相关文献,笔者从构建的信息源和方法两个角度进行梳理。

2.1.1 生物医学领域本体构建信息源

生物医学领域本体构建的信息源主要有两种类型:一是叙词表、分类表、主题词表等受控词表,这类信息源属于结构化信息源,语义结构清晰,概念与概念关系成熟、准确,可以直接作为本体构建的语义基础;二是电子病历、医学文献、在线健康信息等非结构化信息源,数量庞大,包含更加丰富和多样的术语、概念与概念关系,能够体现用户健康信息需求。目前,后者是现今本体构建信息源的研究重点。

(1) 基于结构化信息源的生物医学领域本体构建。结构化信息源中,研究者使用最多的是UMLS、SNOMED CT、MeSH这3种信息源。①基于UMLS的本体构建。UMLS语义网络是本体构建中使用最多的信息源,常被作为顶层本体用于生物医学领域本体构建的基础。李亚子等^[6]基于UMLS语义网络构建症状、检查、药物、医疗法规等子领域本体,并将其与UMLS语义类型映射,借助语义类型之间的关系实现各个子领域本体的关联。米杨等^[7]充分利用UMLS语义网络中的语义关系,构建具有通用语义属性的医学文献本体。曹锦丹等^[8]基于UMLS语义网络中的语义属性关系构建

具有上层语义特征的领域本体(国家基本药物本体)。

②基于SNOMED CT与MeSH的本体构建。Shepherd等^[9]基于SNOMED CT构建了一个本体并将其作为边界对象,以解决照顾慢性病患者多学科卫生保健小组成员之间的语义互操作鸿沟。Naskar等^[1]基于MeSH、SNOMED CT等标准词表使用分面方法构建了人类神经系统(Human nervous system, HNS)疾病本体,包含疾病、诊断、症状、副作用等概念类型。牟冬梅等^[10]结合SNOMED CT和概念格构建了甲状腺疾病本体。李晓瑛等^[11]在借鉴已有疾病本体描述框架的基础上,复用SNOMED CT、UMLS等权威医学知识组织系统中肿瘤相关概念及内容结构,从肿瘤(名称)、病因、诊断、治疗4个维度构建肿瘤本体。

(2)基于非结构化信息源的生物医学领域本体构建。相较于受控词表,电子病历、在线健康信息、医学文献等非结构化信息源具有更加丰富的待挖掘的语义信息,成为现今本体构建的重要信息源。①基于电子病历的本体构建。Almeida等^[12]以电子健康记录和医学信息系统中的医学信息为基础构建了产科和新生儿本体

(Obstetric and Neonatal Ontology, OntoNeo),目的在于组织专门的医学知识、统一表示相关信息。陆泉等^[13]利用电子病历大数据与疾病知识的联系,构建了基于扩展疾病本体DO的电子病历大数据组织模型框架,并以维基百科、MIMIC为例展开DO扩展与电子病历大数据知识组织实验。②基于在线健康信息的本体构建。Choi^[14]通过对社交网站收集的社会化标签进行标签语义分析,提出基于社会化标签构建用户健康本体的框架,以更好地满足健康用户需求,改进用户访问与检索。Zhitomirsky-Geffet等^[15]基于用户对知识领域的各种观点均感兴趣,通过两阶段方法建立非专家用户和群体智慧协作的多角度本体,并以饮食对健康的影响为例进行实证。Clunis^[16]基于网络食谱开发了用于管理高血压患者饮食的本体,该本体描述了食谱、食物中的营养成分、营养成分与处方药的相互作用、疾病与一般健康的关系,为高血压等慢性疾病患者的治疗提供支持。熊回香等^[17]利用网络社区用户的群体需求特征,复用和优化已有的通用疾病本体模型,构建面向用户的慢病知识服务模型,并构建了冠心病本体及相应慢病知识服务平台。③基于医学文献的本体构建。周利琴等^[18]采用人工和自动相结合的方式从非结构化数据源《中国高血压防治指南2010》中抽取知识元素,并将其融合到

DO疾病本体中的高血压本体,对该高血压本体进行扩充,形成可以解决具体领域问题的知识库。于凡等^[19]利用本体技术从《中国2型糖尿病防治指南(2013年版)》中抽取概念并建立语义关联,通过七步法和骨架法构建糖尿病本体库。

2.1.2 生物医学领域本体构建方法

生物医学领域本体的构建方法主要包括人工构建、自动构建、半自动构建3种类型。①人工构建是指通过Protégé、Kano、Chimaera、WebOnto等编辑器或构建工具构建本体,具有操作方式简单、概念与概念关系准确、工作量大等特点。②自动构建是指借助机器学习、数据挖掘、人工智能等技术,通过自然语言规则以及统计分析明确并构建概念以及概念间关系^[20],可大幅减少工作量,在抽取大规模网络数据中具有巨大优势。③半自动构建是人工方式与自动方式的结合,首先借助自然语言处理、数据挖掘等自动化方法,从语料中抽取概念与概念关系,然后人工校验抽取结果^[20]。

(1)基于人工方式的生物医学领域本体构建。Kostyuk等^[21]通过专家手工标注274篇摘要的方式,鉴定了30种出现在自闭症医学文献中的语言障碍术语,并将其组织成自闭症语言障碍本体。杜志银等^[20]提出以临床检验诊断应用为目的,基于人工方式与现有领域资源复用构建临床检验诊断学领域双语本体。Naskar等^[1]使用分面方法构建了HNS疾病本体。

(2)基于自动或半自动方式的生物医学领域本体构建。Rios-Alvarado等^[22]通过层次关系发现和公理提取相结合的**本体学习方法发现医学领域的层次关系和公理抽取。Peng等^[23]提出基于网络资源的本体学习方法构建医学本体,首先根据医学词表抽取术语和分类关系来构建轻量级本体,然后从网络资源中自动学习非分类关系,最后将轻量级本体集成为大型医学本体。洪亮等^[24]提出通过推理-复用方法(半自动方法)来构建医学本体,首先初始化领域知识和数据,确保本体知识结构的准确性;然后使用基于本体和规则的知识推理来扩展本体规模,并基于该方法首次构建了脑区-自闭症本体。唐晓波等^[25]通过基于形式概念分析的本体构建方法,包括数据获取、特征词抽取、形式概念分析和OWL本体描述4个模块完成本体构建。

2.2 生物医学领域本体的整合

现有的生物医学领域本体大多单一存在,并只针对某一方面知识进行概括。面对生物医学领域中问题的多样性和复杂性,多个本体融合能够有效促进领域知识的共享和表达,提供多维度解决问题的视角,并能够支持领域知识的重用和互操作性。本体整合须对本体异构特征进行分析,本体整合的主要方式为映射。

国外方面,Raje等^[26]分析了DO和SNOMED CT在疾病覆盖范围与层次结构上的差异,有助于这两种本体之间的互操作。Finke等^[27]使用本体映射的方法整合了放射学伽玛斯本体(The Radiology Gamuts Ontology, RGO)、DO与HPO,映射结果可被用于支持自动诊断推理、数据挖掘和知识发现。Kahn等^[28]使用半自动字符串匹配的方法实现了孤儿罕见病本体(the Orphanet Rare Disease Ontology, ORDO)与RGO的术语匹配,扩展了可用于放射鉴别诊断的术语和知识范围。

国内方面,王丽伟等^[29]提出多领域本体映射与聚类理论模型,并以该模型为指导,选取药物领域本体RxNorm与NDF-RT(美国国家药物文件——参考术语)进行映射实例研究,提出RxNorm与NDF-RT两个领域本体之间映射及信息分类与聚合的新方法。曹春萍等^[30]通过分析GO、DO和HPO这3个本体的结构和内容,回溯本体构建过程,从构建数据源中挖掘不同本体间存在的关联关系,从而实现本体融合,促进从基因分子水平上剖析疾病的产生以及疾病在体征上的表达等因果串联性研究。程亮^[31]根据疾病术语之间的同义词关系以及本体中存在的集合包含关系,将疾病数据库中的疾病术语分别以同义词映射以及推理映射的方式映射到疾病本体的术语上,从而整合了不同疾病数据库中的疾病术语。米杨^[32]基于顶层本体整合构建了医学领域本体语义标注系统模型,探讨本体整合对于语义标注效果的促进作用。

2.3 生物医学领域本体的扩展

传统生物医学领域本体,如UMLS、SNOMED CT、DO、GO等,虽然具有较高的准确性、权威性与规范性,但多为专业术语,在以非专业词汇为主的面向用户的应用中不能提供充分支持,因此,需要利用在线健康信息、电子病历、医学文献等资源来扩展传统生物医学领域本体中的术语、概念、概念关系。

Muresan等^[33]从面向消费者的文本和网络文档中自动抽取定义,通过归纳语义语法的方法将自然语言表达的定义映射到术语知识,以实现从文本中构建消费者健康术语。Cuzzola等^[34]使用简单知识组织系统(SKOS)和资源描述框架(RDFS)中的本体关系将UMLS概念映射到DBPedia资源,以缩小专业人士与公众使用的健康术语之间的差距,为生物医学领域的知识库构建提供基础。李晓瑛等^[11]探讨了一种基于生物医学文献主题标引词的语义关系发现方法,用于丰富肿瘤本体类间关系及扩充肿瘤本体知识库。

3 生物医学领域本体的相关实践

生物医学领域本体包含规范概念与语义信息,能够解决相关概念的规范描述,提高医学信息的检索效果,主要应用于医学信息描述、查询扩展、知识库构建、术语相似度计算。

(1) 医学信息描述。Liu等^[35]利用基于DO疾病本体的注释来描绘生物医学研究活动,为使用现有本体自动分析疾病研究活动提供了证明。Albright等^[36]利用UMLS中的语义信息,对带有127 606个标记的临床叙事语料库进行人工标注。Martinez-Costa等^[37]基于本体实现相同或相似临床信息的异构语法表示,从而提高电子健康记录(EHRs)的语义互操作。米杨等^[38]利用Protégé、GATE等工具整合中文鼻部炎症疾病知识本体和国家基本药物知识本体,实现对电子病历的标注。

(2) 查询扩展。Azcarate等^[39]使用医学主题词表MeSH的层级结构进行查询扩展,以提高医学图像检索系统的检索效率。Yim等^[40]利用生物医学本体和丰富的语言特征对放射学报告中肿瘤事件属性进行分类。Yu等^[41]将基因组学研究中的重要概念与UMLS语义网络整合。

(3) 知识库构建。Chen等^[42]使用UMLS结构构建医学图像库,并基于标准医学本体中的术语来标注图像,如解剖学基础模型(FMA)、国际疾病分类第9版(ICD9)、RxNorm等,为消费者建立大规模的健康图像库。咸珂^[43]采用基于本体的信息抽取算法从在线健康问诊内容中抽取疾病症状、治疗方案、所需检查等信息,并形成结构化的健康知识库。巩沐歌等^[44]构建了基于本体的具有智能推理功能的高血压疾病知识库,为高血压诊断专家系统建立可共享、可复用的诊断知识库系统提供参考。

(4) 术语相似度计算。邵玉凯^[45]利用HPO研究基因和表型、疾病和表型的相似性关系,进而预测病人的致病基因和疾病,提出一种基于HPO中通路的相似性计算方法。范雪雪等^[46]依据SNOMED CT和MeSH两个医学本体的层级结构和语义关系,提取概念术语的深度、距离等语义参数,并用概念密度对其加权得到深度系数和距离系数,构造相似度函数进行术语相似度计算。邱实^[47]将生物医学领域本体与疾病实验数据建立映射关联,从疾病实验数据中发现更多概念与概念之间的相似关系。

4 生物医学领域本体的发展趋势

基于上述不同方面中相关实践的归纳总结,生物医学领域本体的未来发展趋势如下。

(1) 生物医学领域本体的应用范围逐步扩大。除传统的医学信息描述、查询扩展之外,基于本体的生物医学领域知识库构建以及术语相似度计算方面的应用在不断增加。

(2) 生物医学领域本体的数据来源类型从传统结构化信息源逐渐转向非结构化信息源。相较于传统的受控词表等结构化信息源,电子病历、在线健康信息、医学文献等非结构化信息源包含更加丰富、更符合用户需求的语义信息,成为传统生物医学领域本体扩展及新型生物医学领域本体构建的重要数据来源,能够更好地适用于面向用户的健康应用中。

(3) 生物医学领域本体的构建类型与构建方法分别向精准化、自动化发展。特定生物医学子领域本体是现今本体构建热点,机器学习、自然语言处理等本体学习方法被广泛应用于从大量生物医学领域数据中抽取用于构建本体的概念与概念关系。

(4) 多个本体融合是解决复杂生物医学领域问题的重要方法。生物医学问题往往关联疾病的多个方面,以具体生物医学问题为核心,关联与其相关的多个异构本体,实现本体重用,并整合相关领域知识。

5 结论

本文首先从本体概述、创建单位、本体类型、本体规模、本体结构5个方面归纳总结了国内外具有代表性的生物医学领域本体。同时,通过对2010—2019年国内外生物医学领域本体的文献调研,从构建、整合与扩展

3个方面总结了相关研究,以及生物医学领域本体在医学信息描述、查询扩展、知识库构建、术语相似度计算等方面的应用实践,并分析了生物医学领域本体的未来发展趋势。构建方面,笔者分别从构建信息源与构建方法两个角度归纳,构建信息源包括UMLS、SNOMED CT、MeSH等结构化信息源,以及电子病历、在线健康信息、医学文献等非结构化信息源;构建方法包括利用Protégé等人工方式构建,以及形式概念分析、深度学习等自动或半自动方式构建。整合方面,主要包括对不同本体异构特征的分析以及基于映射的本体整合。扩展方面,利用在线健康信息、电子病历、医学文献等资源来扩展传统生物医学领域本体中的术语、概念与概念关系。

参考文献

- [1] NASKAR D, DAS S. HNS ontology using faceted approach [J]. Knowledge Organization, 2019, 46 (3): 187-198.
- [2] 陈焱,姜慧敏. 医学领域本体研究现状 [J]. 情报科学, 2006 (10): 149-152.
- [3] 吴明智,崔雷. 生物医学相关的本体研究现状 [J]. 医学信息学杂志, 2009, 30 (7): 41-44.
- [4] 张庆,吕少妮,轩扬. 本体在生物医学领域中应用研究热点分析 [J]. 医学信息学杂志, 2019, 40 (1): 63-67.
- [5] 吴正荆,黄薇,牟冬梅,等. 生物医学领域本体开发项目比较研究 [J]. 中华医学图书情报杂志, 2010 (5): 20-23.
- [6] 李亚子,钱庆,刘峥,等. 基于UMLS的疾病知识整合框架研究 [J]. 现代图书情报技术, 2011 (2): 34-41.
- [7] 米杨,曹锦丹. 利用UMLS语义网络构建医学文献本体的实证探索 [J]. 图书馆学研究, 2012 (7): 55-60.
- [8] 曹锦丹,米杨,周喜龙. 基于顶级本体上层语义的领域本体构建研究 [J]. 情报科学, 2014, 32 (9): 142-146.
- [9] SHEPHERD M, SAMPALLI T. Ontology as boundary object [C] // Categories, Contexts and Relations in Knowledge Organization: Proceedings of 12th International ISKO Conference 6-9 August 2012 Mysore, India. Würzburg, Ergon Verlag, 2012: 131-137.
- [10] 牟冬梅,张艳侠,黄丽丽,等. 基于SNOMED CT和FCA的医学领域本体构建研究 [J]. 情报学报, 2013, 32 (6): 653-662.
- [11] 李晓瑛,李丹亚,夏光辉,等. 肿瘤本体构建研究 [J]. 数字图书馆论坛, 2015 (8): 37-42.
- [12] ALMEIDA M B, FARINELLI F. Ontologies for the representation of electronic medical records: the obstetric

- and neonatal ontology [J]. *Journal of the Association for Information Science and Technology*, 2017, 68 (11): 2529-2542.
- [13] 陆泉, 江超, 陈静. 基于扩展疾病本体的电子病历大数据组织研究 [J]. *图书情报知识*, 2019, 187 (1): 111-120.
- [14] CHOI Y. Supporting better treatments for meeting health consumers' needs: extracting semantics in social data for representing a consumer health ontology [J]. *Information Research: An International Electronic Journal*, 2016, 21 (4).
- [15] ZHITOMIRSKY-GEFFET M, EREZ E S, JUDIT B I. Toward multiviewpoint ontology construction by collaboration of non-experts and crowdsourcing: the case of the effect of diet on health [J]. *Journal of the Association for Information Science and Technology*, 2017, 68 (3): 681-694.
- [16] CLUNIS J S. Designing an ontology for managing the diets of hypertensive individuals [J]. *International Journal on Digital Libraries*, 2019, 20 (3): 269-284.
- [17] 熊回香, 代沁泉, 梅潇. 面向在线医疗社区的慢病知识服务模型构建 [J]. *情报理论与实践*, 2020, 43 (6): 123-130.
- [18] 周利琴, 范昊, 潘建鹏. 网络大数据中的知识融合框架研究 [J]. *情报杂志*, 2018, 37 (1): 145-150, 197.
- [19] 于凡, 雷行云, 高星, 等. 基于临床指南的糖尿病本体构建及语义检索模型设计 [J]. *医学信息学杂志*, 2018, 39 (5): 45-50.
- [20] 杜志银, 刘芳, 胡虹, 等. 临床检验诊断学双语领域本体构建方法研究 [J]. *情报探索*, 2010 (6): 18-20.
- [21] KOSTYUK N, ISOKPEHI R D, RAJNARAYANAN R V, et al. Areas of language impairment in autism [J]. *Autism Insights*, 2010 (2): 31-38.
- [22] RIOS-ALVARADO A B, LOPEZAREVALO I, TELLOLEAL E, et al. An approach for learning expressive ontologies in medical domain [J]. *Journal of Medical Systems*, 2015, 39 (8): 1-15.
- [23] PENG J, DU Y, CHEN Y, et al. Medical ontology learning based on web resources [C] // 2015 12th Web Information System and Application Conference (WISA), IEEE, 2015: 116-119.
- [24] 洪亮, 石晓月. 医学本体构建方法研究——以脑区与自闭症为例 [J]. *信息资源管理学报*, 2020, 10 (2): 80-90.
- [25] 唐晓波, 伍星蓉. 基于形式概念分析的面向药物适应症的领域本体构建方法 [J]. *情报杂志*, 2017 (12): 159-165.
- [26] RAJE S, BODENREIDER O. Interoperability of disease concepts in clinical and research ontologies: contrasting coverage and structure in the disease ontology and SNOMED CT [J]. *Studies in Health Technology & Informatics*, 2017, 245: 925-929.
- [27] FINKE M T, FILICE R W, KAHN C E, et al. Integrating ontologies of human diseases, phenotypes, and radiological diagnosis [J]. *Journal of the American Medical Informatics Association*, 2019, 26 (2): 149-154.
- [28] KAHN C E. Integrating ontologies of rare diseases and radiological diagnosis [J]. *Journal of the American Medical Informatics Association*, 2015, 22 (6): 1164-1168.
- [29] 王丽伟, 王伟, 高玉堂, 等. 领域本体映射的语义互联方法研究——以药物本体为例 [J]. *图书情报工作*, 2013, 57 (17): 21-25, 33.
- [30] 曹春萍, 张政. 基于本体参考数据的生物医学本体融合模型研究 [J]. *计算机应用与软件*, 2018, 35 (10): 38-44.
- [31] 程亮. 基于本体的疾病数据整合与挖掘方法研究 [D]. 哈尔滨: 哈尔滨工业大学, 2014.
- [32] 米杨. 基于顶级本体整合的医学领域语义标注研究 [D]. 长春: 吉林大学, 2012.
- [33] MURESAN S, KLAVANS J L. Inducing terminologies from text: a case study for the consumer health domain [J]. *Journal of the American Society for Information Science and Technology*, 2013, 64 (4): 727-744.
- [34] CUZZOLA J, BAGHERI E, JOVANOVIĆ J. UMLS to DBpedia link discovery through circular resolution [J]. *Journal of the American Medical Informatics Association*, 2018, 25 (7): 819-826.
- [35] LIU Y, COULET A, LEPENDU P, et al. Using ontology-based annotation to profile disease research [J]. *Journal of the American Medical Informatics Association*, 2012, 19 (e1): e177-e186.
- [36] ALBRIGHT D, LANFRANCHI A, FREDRIKSEN A, et al. Towards comprehensive syntactic and semantic annotations of the clinical narrative [J]. *Journal of the American Medical Informatics Association*, 2013, 20 (5): 922-930.
- [37] MARTINEZ-COSTA C, CORNET R, KARLSSON D, et al. Semantic enrichment of clinical models towards semantic interoperability: the heart failure summary use case [J]. *Journal of the American Medical Informatics Association*, 2015, 22 (3): 565-576.
- [38] 米杨, 曹锦丹. 顶级本体统控的多本体语义标注实证研究 [J]. *数据分析与知识发现*, 2012 (9): 36-41.
- [39] AZCARATE M C, VAZQUEZ J M, LOPEZ M J, et al.

- Improving image retrieval effectiveness via query expansion using MeSH hierarchical structure [J]. Journal of the American Medical Informatics Association, 2013, 20 (6) : 1014-1020.
- [40] YIM W, KWAN S W, YETISGEN M, et al. Classifying tumor event attributes in radiology reports [J]. Journal of the Association for Information Science and Technology, 2017, 68 (11) : 2662-2674.
- [41] YU H, FRIEDMAN C, RHZETSKY A, et al. Representing genomic knowledge in the UMLS semantic network [C] // Proceedings Annual Symposium. 1999, 6 (1) : 181.
- [42] CHEN Y, REN X, ZHANG G Q, et al. Ontology-guided organ detection to retrieve web images of disease manifestation: towards the construction of a consumer-based health image library [J]. Journal of the American Medical Informatics Association, 2013, 20 (6) : 1076-1081.
- [43] 咸珂. 基于本体的健康知识库自动构建方法研究 [D]. 哈尔滨: 哈尔滨工业大学, 2016.
- [44] 巩沐歌, 温有奎. 基于本体的高血压疾病诊断知识库 [J]. 情报杂志, 2010, 29 (S1) : 169-172.
- [45] 邵玉凯. 基于人类表型本体的基因和疾病关联关系分析方法研究 [D]. 哈尔滨: 哈尔滨工业大学, 2015.
- [46] 范雪雪, 王志荣, 徐晤, 等. 基于医学本体的术语相似度算法研究 [J]. 现代图书情报技术, 2015, 31 (12) : 57-64.
- [47] 邱实. 基于领域本体的生物医学本体匹配算法研究 [D]. 哈尔滨: 哈尔滨工业大学, 2015.

作者简介

赵洁, 女, 1990年生, 博士, 讲师, 研究方向: 知识组织, E-mail: zhaojie_shuique@163.com。
司莉, 女, 1965年生, 博士, 教授, 研究方向: 知识组织、科学数据管理。

Research and Practice Progress of Biomedical Domain Ontology at Home and Abroad

ZHAO Jie¹ SI Li²

(1. School of Information, Shanxi University of Finance & Economics, Taiyuan 030006, China; 2. Center for the Studies of Information Resources of Wuhan University, Wuhan 430072, China)

Abstract: By means of literature review, the paper summarizes the ontological literatures in the biomedical domain at home and abroad in the past ten years (from 2010 to 2019). This paper briefly introduces the important biomedical domain ontology at home and abroad, summarizes the relevant research of biomedical domain ontology from three aspects, which are construction, integration, and extension, and its related practice in medical information description, query expansion, knowledge base construction, term similarity calculation. And the future development trend is prospected in order to provide a reference for the construction and application of biomedical domain ontology in China.

Keywords: Biomedical Domain; Ontology; Ontology Construction; Ontology Integration; Ontology Expansion; Ontology Application

(收稿日期: 2020-05-12)