

黑色素瘤单病种本体构建研究*

康宏宇

(1 中国医学科学院/北京协和医学院医学信息研究所 北京 100020

2 北京理工大学生命学院生物医学工程系 北京 100081)

李 勤

吴 萌 沈 柳 李 姣 侯 丽

(北京理工大学生命学院生物医学工程系
北京 100081)

(中国医学科学院/北京协和医学院医学信息研究所
北京 100020)

〔摘要〕 利用百科知识、临床指南、科技文献、专家经验等多源异构资源进行本体顶层范畴界定,借助 Ontofox、Protégé 等本体工具采集并组织黑色素瘤疾病、分子异常、症状、解剖结构、药物、干预与操作、并发症等顶层范畴的概念和术语,构建黑色素瘤单病种本体并对其评估,开发本体展示和检索工具,为相关领域研究提供参考。

〔关键词〕 单病种;黑色素瘤;本体构建;Ontofox;知识检索

〔中图分类号〕 R-058 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2021.12.007

Study on the Single-disease Ontology Construction of Melanoma KANG Hongyu, 1Institute of Medical Information, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100020, 2Department of Biomedical Engineering, School of Life Science, Beijing Institute of Technology, Beijing 100081, China; LI Qin, Department of Biomedical Engineering, School of Life Science, Beijing Institute of Technology, Beijing 100081, China; WU Meng, SHEN Liu, LI Jiao, HOU Li, Institute of Medical Information, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100020, China

〔Abstract〕 Multi-source heterogeneous resources such as encyclopaedia, clinical guidelines, scientific literature and expert experience are used to define the top-level category of ontology, and ontology tools such as Ontofox and Protégé are used to collect and organize top-level concepts and terms of melanoma diseases, molecular abnormalities, symptoms, anatomical structures, drugs, interventions and procedures, complications, etc. The single-disease ontology of melanoma is constructed and evaluated, the ontology visualization and retrieval tools are developed, and some useful references are provided for ontology construction research.

〔Keywords〕 single-disease; melanoma; ontology construction; Ontofox; knowledge retrieval

〔修回日期〕 2021-09-22

〔作者简介〕 康宏宇,助理研究员,发表论文 12 篇;通讯作者:侯丽,博士,研究员,发表论文 60 余篇。

〔基金项目〕 中央高校基本科研业务费“面向精准用药的药物基因组学知识图谱构建”(项目编号:3332020049);医药卫生专业知识服务系统(项目编号:CKCEST-2020-2021-1-6);新闻出版业科技与标准重点实验室“医学融合出版知识技术重点实验室”。

1 引言

本体是在一个具体学科领域中, 由该学科专业术语及术语间关系构成的一种复杂的知识网络体系, 是对领域知识、概念及其相互关系进行的规范化描述和组织, 进而揭示概念之间更完整的语义关系和事物属性, 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^[1]。随着医学领域大数据的广泛积累和信息技术快速发展, 本体由于其良好的结构化表示信息能力和人机可读性特点成为生物医学领域资源组织和标准化的主要方式, 已经被广泛应用于生物医学数据分析、检索、整合、共享和再利用, 其对知识的有效管理能够支撑疾病早期诊断和预防、临床决策系统、医药卫生知识查询和服务等实践应用环节。本文在本体构建经典方法——七步法的基础上, 根据实际需求进行步骤调整, 以黑色素瘤为切入点, 使用 Ontofox 工具对领域内的通用本体进行复用, 最终构建涵盖疾病、药物、解剖结构、症状、干预与操作等 7 类黑色素瘤发病、诊断、治疗及相关概念类的黑色素瘤单病种本体, 对类与类之间的 34 种语义关系进行定义和丰富。

2 生物医学领域本体构建现状分析

2.1 单病种本体构建进展

2.1.1 概述 生物医学领域本体是以解决生物医学学术语通用性、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^[1]。近年来典型生物医学领域通用本体有基因本体 (Gene Ontology, GO), 疾病本体 (Disease Ontology, DO), 人类表型本体 (Human Phenotype Ontology, HPO) 等。

2.1.2 国内研究情况 糖尿病是全球重要的慢性疾病之一, 中国医学科学院医学信息研究所、重庆医科大学附属第二医院、华中科技大学同济医学院、吉林大学管理学院等多家研究机构利用糖尿病临床指南^[2]、临床经验和专家指南^[3-4]、医学主题

词表 (Medical Subject Headings, MeSH) 等领域通用词表^[5]等对糖尿病病因、临床表现、并发症、药物治疗、诊断方法、生活干预方式等进行综合知识汇总, 进而构建糖尿病本体, 将糖尿病疾病本身及药物、诊断等相关知识进行关联和结构化管理, 部分研究还实现了基于本体的语义推理。随着知识抽取、知识融合技术的发展, 洪亮和石晓月^[6]利用基于本体和规则的知识推理扩展本体规模, 提出“推理-复用本体”的构建方法并构建脑区-自闭症本体。

2.1.3 国外研究情况 国外研究者面向糖尿病^[7-8]、甲状腺^[9]等疾病, 从患者现状、疾病类型、药物、并发症、症状、实验室检测、药物相互作用等类型出发构建单病种本体, 进而更有效地支持临床决策等实践环节。单病种本体的构建能够更加聚焦某一种疾病的预防、诊断、治疗全过程, 帮助临床医生获取规律性认识、筛选出支持临床决策的直接参考证据, 进一步支撑疾病研究过程中的文献挖掘、智能检索及知识推理等应用。

2.2 通用本体构建方法

2.2.1 手工构建方法 相关研究较多, 具有代表性的有七步法、骨架法、SENSUS 法、KACTUS 法、IDEF5 法、METHODOLOGY 法、循环获取法、TOVE 法等。岳丽欣和刘文云^[10]对以上 8 种较为成熟的本体构建方法进行介绍和对比分析; 尚新丽^[11]在以上 8 种方法的基础上增加五步循环法, 对比分析各本体构建方法特点和不足、适用领域和现实指导意义。七步法^[12]由斯坦福大学医院开发, 是一种目前公认的更加适用于领域本体构建的方法。本文对七步法进行调整和完善, 构建黑色素瘤单病种本体。

2.2.2 自动构建方法 本体的自动构建通过机器学习、自然语言处理及人工智能等方法自动从大量自由文本以及机器可以识别的词典、叙词表、数据库等数据源中获取概念及概念间关系, 从而完成本体构建^[13]。其关键步骤包括数据预处理、术语自动抽取、关系自动抽取、合成本体。Mazen A、Mahmood M K 和 Susan S^[14]提出一种基于生物医学领域

关联开放数据 (Linked Open Data, LOD) 的本体生成方法 (LOD - ABOG), LOD - ABOG 首先利用自然语言处理技术基于统一医学语言系统 (Unified Medical Language System, UMLS) 以及关联开放数据进行概念提取, 再使用广度优先搜索 (Breadth - First Search, BFS) 图方法和 Freepal 存储库模式实现关系提取。本体自动构建方法能够处理大规模数据且成本低、效率高, 但由于各类语言表述中存在复杂性, 目前出现了利用人工参与、机器辅助的半自动方法来构建本体。

3 黑色素瘤本体构建 (图 1)

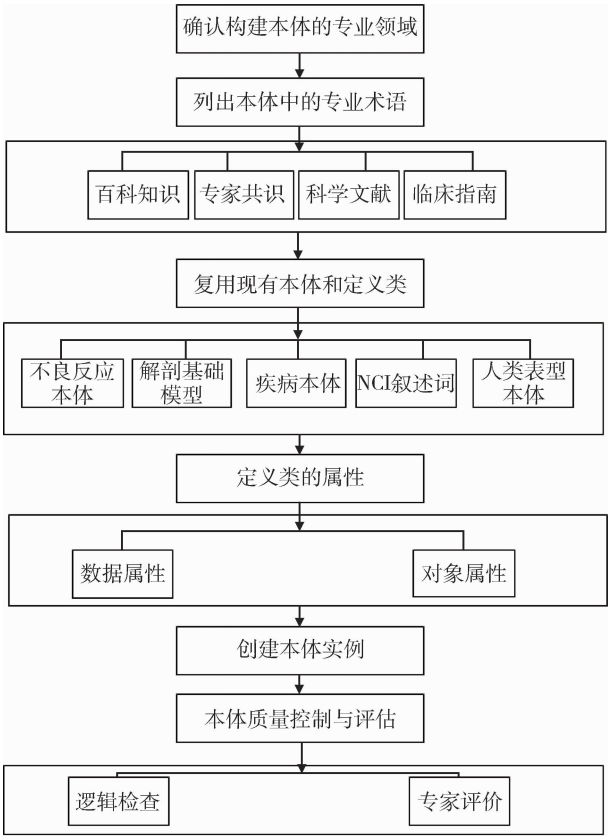


图 1 黑色素瘤单病种本体构建流程

3.1 确定顶层学科范畴

黑色素瘤是皮肤癌中最致命的恶性肿瘤, 其发病率和病死率持续上升, 年增长率为 3% ~ 7%^[15], 是发病率增长最快的恶性肿瘤之一^[16]。通过文献综

述与专家咨询发现黑色素瘤的发病与 DNA 的损伤有密切关系, 其发病多见于皮肤、肢端、黏膜等身体部位, 诊断和治疗手段包括影像学检查、活检、化疗、放疗、免疫治疗、手术治疗、靶向药物治疗等。建立全面的顶层学科范畴能够进一步支持疾病的临床诊疗全流程; 同时作为单病种本体, 其学科范畴还应该涉及该疾病的症状、并发症等内容。因此本研究以黑色素瘤为起点, 将黑色素瘤的顶层学科范畴体系确定为疾病、分子异常、症状、解剖结构、药物、干预与操作、并发症。

3.2 提取关键术语

3.2.1 关键术语提取来源 在确定领域范畴的基础上列出本体关键术语, 便于清晰梳理本体层级结构, 围绕关键术语开展进一步概念扩充和语义关系发现。关键术语提取来源有百科知识^[17-18] 专家共识, 科学文献, 临床指南 (包括中国黑色素瘤诊治指南 (2011 版), 黑色素瘤诊疗规范 (2018 年版), 美国国家综合癌症网络 (National Comprehensive Cancer Network, NCCN) 临床实践指南 (2019. v1), 欧洲肿瘤内科学会 (European Society for Medical Oncology, ESMO) 发布的《皮肤黑色素瘤的诊断、治疗与随访指南》等)。

3.2.2 标注方法 邀请 3 位具有医学背景的研究人员完成关键术语的标注和提取任务。每份资料均由其中两位研究人员分别进行注释; 产生分歧时第 3 位高级标注人员具有最终决定权。标注人员从上述材料中标注相关实体, 通过 Ontobee 数据库 (<http://ontoanimals.bmicc.cn/ontobee/>) 检索该词的准确性, 找到其来源 ID (Source ID) 并记录。

3.2.3 标注规则 标注内容为 7 类顶层范畴; 如果标注实体在 Ontobee 中对应多条记录, 则尽量按照全名称匹配原则匹配; 如果标注实体在 Ontobee 中是作为同义词出现, 则将其对应的优选词作为标注结果, 这个词本身作为同义词; 如果标注实体在 Ontobee 中没有对应的记录 (即没有 Source ID), 则其编码处标记为 -1; 标注具体症状名称时, 注意症状和疾病的区别, 如果有明确的疾病 ID, 优先标注为疾病; 实体标注完成后, 需要进一步核实其在

Ontobee 中涉及的下位类是否跟黑色素瘤诊疗过程相关,若相关也将其下位词纳入标注术语列表中。此外,使用 F 值来衡量实体标注的一致性,结果显示黑色素瘤实体标注的 F 值为 82.26%。最后结合领域权威专家建议进行领域范畴和关键术语确定。

3.3 复用现有本体、定义类

3.3.1 构建方法 利用自顶向下和自底向上相结合的方法构建黑色素瘤本体层级关系。首先确立顶层概念,对照领域本体将层级关系逐层细化;使用 Ontofox 工具,在顶层概念和重要术语的区间范围内,提取所有中间术语及其定义和属性。

3.3.2 Ontofox 基于网页开发和使用的系统,可以快速提取已知本体的术语及关系,并在新建本体中再利用。其允许用户输入前期制定好的重要术语并直接从源本体获取定义、属性、同义词、注释等术语相关内容,同时使用本体语言 (Ontology Web Language, OWL) 格式保存本体复用结果。

3.3.3 复用本体 并发症相关术语来自不良反应本体 (Ontology of Adverse Events, OAE),解剖结构相关术语来自解剖基础模型 (Foundational Model of Anatomy, FMA),疾病相关术语来自疾病本体,药物、诊断、治疗、手术、分子异常来自 NCI 叙述词 (NCI Thesaurus OBO Edition, NCI),症状相关术语来自人类表型本体。

3.4 定义类属性和属性取值范围

3.4.1 类的数据属性 建立类的数据属性能够较为全面地描述类,并且为开发基于本体的检索、查询等应用奠定基础。结合本文采集的数据特性将本体数据属性定义为:来源编码 (Source ID)、定义 (Textual_definition)、NCI 优选词 (Display_Name)、同义词 (Has_Exact_Synonym、Has_Broad_Synonym、Has_Narrow_Synonym)、其他数据源中的唯一标识 (包括 FDA_UNII_code、UMLS_CUI、CAS_Registry、CHEBI_ID、SNP_ID 等)。

3.4.2 类的对象属性 对象属性包括两种:类与子类之间的上下位关系以及人工定义的类与类之间的语义关系。在此重点介绍类与类之间的语义关

系。本研究围绕上文定义的 7 大类顶层范畴建立 34 种语义关系,构建来源和方式包括 NCI、RO (Relationship Ontology)、OAE、HPO 和自定义。其中 Variation_Is_Biomarker_Of、Variation_Is_Biomarker - diagnosis_Of、Disease_Has_Symptom_Of、Drug_Syn - ergized_By_Associated_Drug、Disease_Has_Associated _Anatomic_Site、Pharmacogenomic_Effects_Disease、Is_Target_Drug_Of_Gene_Mutation、Disease_Treated _by_Drug、Is_Substance_That_Treats_Disease、Dis - ease_Has_Molecular_Abnormality 10 类关系的取值范围在 1 级类的基础上进行细化,覆盖到本体的 2 级或 3 级类,为后续研究知识推理等应用奠定基础,例如比美替尼 (Binimetinib) 是 BRAF 基因突变 (BRAF Gene Mutation) 的靶向治疗药物 (Is Target Drug Of Gene Mutation),肠胃系统疾病 (Gastroin - testinal System Disease) 的相关解剖部位 (Disease Has Associated Anatomic Site) 包括食道、大肠和小肠 (Esophagus, Large Intestine, Small Intestine)。

3.5 创建实例

本体构建的最后一步是添加实例。实例是概念的具体化,继承概念的属性。在本研究中考考虑到基因突变相关术语的分布存在扁平化和具体化特性,所以在黑色素瘤本体的分子异常 (Molecular Abnormality) 分支中设置了实例,如 BRAF NM _ 004333.4: c.1798G > A、BRAF NP _ 004324.2: p.V600E 等。同时在术语量较多、体系较为复杂的解剖结构 (Anatomical Structure)、干预与操作 (Intervention or Procedure) 分支中也设置了实例,如 Right big toe、Left big toe、Right second toe、Axillary Lymph Node Biopsy 等。

4 结果与分析

4.1 本体统计

统计结果显示本研究构建的黑色素瘤本体共包含疾病 (Disease)、分子异常 (Molecular Abnormality)、症状 (Phenotypic Abnormality)、解剖结构 (Anatomical Structure)、药物 (Drug)、干预与操作

(Intervention or Procedure)、并发症 (Complication) 7 个本体大类, 最深层级关系 11 层, 共包含概念 545 个, 术语 1 471 个, 见表 1。

表 1 黑色素瘤本体数据来源及统计

类型	复用本体	层级数量	概念数量	术语数量
疾病	DO	11	81	228
分子异常	NCh	8	29	202
症状	HPO	8	114	222
解剖结构	FMA	9	146	337
药物	NCh	5	19	128
干预与操作	NCh	8	105	303
并发症	OAE	7	51	51
合计	-	-	545	1 471

4.2 质量控制和本体评估

4.2.1 基本方法 目前本体评估方法包括黄金标准法、应用评价法、数据驱动法、基于指标法和专家评价法。临床诊断对医学领域本体的可信度和准确度提出了更高要求。质量控制和评估并不是本体构建的最后一步, 而应该是贯穿在本体建设的整个生命周期^[19]。黑色素瘤本体构建过程中, 在重要术语抽取和本体构建完成后均进行了质量控制和本体评估, 采用逻辑检查法、专家评价法。

4.2.2 术语抽取的正确性和适用性 邀请两位具有医学背景的研究人员对其进行评估。当产生分歧时会有临床专家作为第三方给出相应建议。本体构

建完成后通过两轮自动检查和人工校对, 删除重复术语 1 处、规避术语的回环错误 0 处, 校正异常字符和拼写错误 0 处等。其中重复术语错误为: Phenotypic Abnormality 分支下存在如下两个层级关系 “Phenotypic Abnormality→Bone Pain” 和 “Phenotypic Abnormality→Pain→Bone Pain” 且出现的概念 “Bone Pain” 其所有属性均相同, 最终删除第 1 条层级关系, 保留第 2 条。此外还通过专家评审的方式进一步保证本体的专业性和正确性, 具体包括分类检查 (即从专业角度判断术语所属类别及本体层级关系复用是否恰当) 和内容检查 (即术语定义和同义词是否正确), 根据专家建议对黑色素瘤本体层级关系设置等内容进行修正和完善。

4.2.3 3 个单病种本体结构性指标数据及概念数对比 考虑到黑色素瘤本体为单病种本体, 涉及概念为疾病诊断、治疗过程中的相关术语, 本研究调研到具有相似本体构建出发点的单病种本体——前列腺癌本体 (Prostate Cancer Ontology, PCAO)^[20]、甲状腺癌本体 (Thyroid Cancer Ontology, TCO)^[21] 并进行本体结构性指标数据和概念数比对, 见表 2、表 3。结构性指标数据显示, 黑色素瘤本体较其他单病种本体设置了更为丰富的属性和实例, 分类均匀且对 65% 以上的概念进行了定义, 但层级关系设置尚有待优化。从概念数量上看, 黑色素瘤本体在疾病、症状、解剖结构、并发症、药物治疗等类下, 概念较为丰富; 但干预与操作概念较少且缺乏患者信息和生活方式相关概念设置。以上内容会在后续工作中继续补充和完善。

表 2 3 个单病种本体结构性指标数据

评价指标	类	实例	属性	最大深度	最多子节点数	平均子节点数	唯一子节点的类	大于 25 个子节点的类	没有定义的类
MELO	545	31	34	11	16	2	39	0	198
PCAO	636	0	0	7	53	4	19	2	636
TCO	578	0	3	8	29	4	9	1	20

表 3 3 个单病种本体概念数对比

评价指标	疾病	分子异常	症状	解剖结构	药物	干预与操作	并发症	患者信息	生活方式	病理学特征
MELO	81	29	114	146	19	105	51	-	-	-
PCAO	59	58	27	-	-	339	-	7	146	-
TCO	65	21	38	36	-	321	-	8	51	38

4.3 模型构建

基于黑色素瘤本体涉及的 7 类概念和 34 种语义

关系构建黑色素瘤临床诊疗知识模型, 见图 2。知识模型更加直观地展示了与黑色素瘤病因、诊断、治疗等临床环节相关的网络结构。

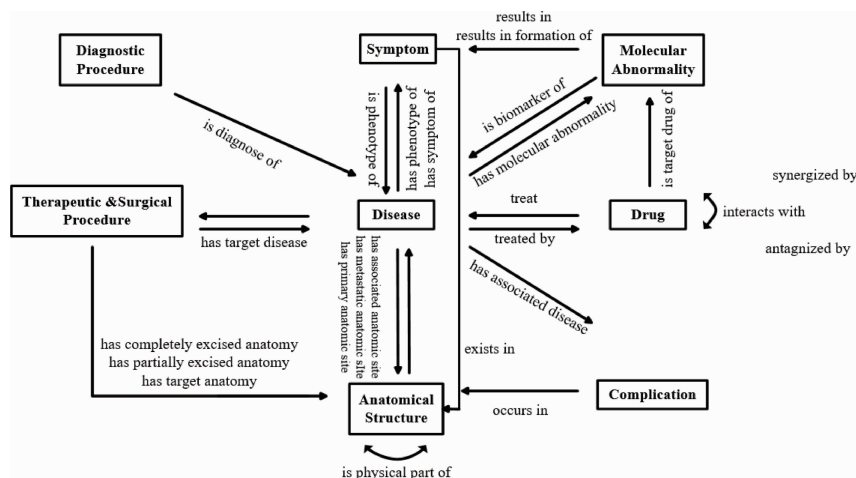


图2 黑色素瘤临床诊疗知识模型

4.4 基于本体的知识展示

在现有研究中, 本体应用方向包括文献挖掘、信息检索、临床决策支持、医疗问答等。为进一步发挥黑色素瘤单病种本体效用, 本研究开发黑色素瘤本体展示和检索工具, 并在中国工程科技知识中心“医药卫生专业知识服务系统”(http://med.ckcest.cn/open-api-ontology.html)中进行数据发布, 以黑色素瘤单病种本体为底层数据, 提供基于本体的检索和浏览服务, 实现本体的可视化展示。用户可以通过检索关键词来找到本体中相应术语, 确定术语在本体中的位置及上下位关系, 浏览术语的对象属性, 为用户了解不同术语层级关系及同义词提供了便利, 为临床科研人员开展基于黑色素瘤相关的知识挖掘与研究提供数据和工具支撑。

5 结论

领域本体是领域概念及其关系的一种形式化描述, 由于具有显著的领域特性, 能够更加合理和有效地进行知识表示。在医学领域, 单病种本体更具有针对性, 能够完整且具体地围绕某一种疾病开展知识组织和关系发现, 在临床实践环节具有更多真

实的应用场景和应用价值。本研究充分利用多来源知识资源, 包括百科知识、专家共识、科学文献、临床指南等, 以知识组织构建为出发点进行理论与技术的探索实践, 即借助 Ontofox、Protégé 等本体复用和本体编辑工具, 在传统七步法的基础上对本体构建步骤进行调整和完善, 围绕黑色素瘤疾病梳理出疾病、药物、症状、干预与操作等 7 大类知识概念, 定义概念与概念间的 34 种语义关系, 完成黑色素瘤临床知识本体构建研究。该本体是将多源异构的疾病知识与本体构建技术相结合构建起来的知识组织体系, 实现了百科、文献、指南等静态文本形式资源和口口相传的专家经验向规范化、标准化、结构化的知识组织体系转换, 能够支持医学知识检索系统实现可查询、可推理, 辅助临床医生制定更细致、更可靠的诊疗方案, 也进一步提高了临床指南等文本资源在疾病防治和诊疗过程中的利用率。另外与常见的领域本体构建研究相比, 本研究介绍了一种本体复用工具——Ontofox, 实现了对 NCI、HPO、OAE 等已有领域本体的快速复用, 大幅提升本体构建效率和准确性。但目前该项研究工作存在一定局限性, 主要有以下几个方面: 目前黑色素瘤本体构建过程由手动实现, 工作量较大, 后续将会对本体自动构建方法进行探索; 对黑色素瘤

本体进行了技术层面的质量评估,但考虑到临床实践环节对数据真实性、信息准确性、响应及时性需求,该本体对临床的辅助程度仍需要对其内容全面性和真实世界可应用性进行验证;同时还需要拓宽黑色素瘤本体的应用场景,将本体嵌入临床决策支持系统及知识发现与推荐系统中,为医生、患者提供更具象化、更深入的知识服务。

参考文献

- 1 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.
- 2 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.
- 3 张莉, 王玉廷. 基于 Protege 的糖尿病本体构建 [J]. 科学咨询 (教育科研), 2020 (3): 9-11.
- 4 刘智锋, 夏晨曦, 黄梨, 等. 糖尿病领域本体构建与语义推理实现 [J]. 中华医学图书情报杂志, 2017, 26 (9): 7-11.
- 5 刘何心. 糖尿病本体的构建与检索研究 [D]. 长春: 吉林大学, 2014.
- 6 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.
- 7 El-Sappagh S, Ali F. DDO: a diabetes mellitus diagnosis ontology [J]. Applied Informatics, 2016, 3 (1): 5.
- 8 El-Sappagh S, Kwak D, Ali F, et al. DMTO: a realistic ontology for standard diabetes mellitus treatment [J]. Journal of Biomedical Semantics, 2018, 9 (1): 8.
- 9 Özgü C, Emine S, Murat O, et al. Implementing Thyroid Ontology for Diagnosis and Care [C]. Barcelona: The Sixth International Conference on Global Health Challenges, 2017.

- 10 岳丽欣, 刘文云. 国内外领域本体构建方法的比较研究 [J]. 情报理论与实践, 2016, 39 (8): 119-125.
- 11 尚新丽. 国外本体构建方法比较分析 [J]. 图书情报工作, 2012, 56 (4): 116-119.
- 12 Noy N F. Ontology Development 101: A Guide to Creating Your First Ontology: Knowledge Systems Laboratory, Stanford University [EB/OL]. [2021-05-28]. <http://www.ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness.pdf>.
- 13 王向前, 桂冬冬, 李慧宗. 面向文本的本体自动构建研究综述 [J]. 图书馆理论与实践, 2019 (4): 45-50.
- 14 Mazen A, Mahmood M K, Susan S. Linked Open Data-based Framework for Automatic Biomedical Ontology Generation [J]. BMC Bioinformatics, 2018, 19 (1): 319-332.
- 15 《临床肿瘤学杂志》编辑部. 中国黑色素瘤诊治指南 (2011 版) [J]. 临床肿瘤学杂志, 2012, 17 (2): 159-171.
- 16 赵辨. 中国临床皮肤病学 [M]. 南京: 江苏科技技术出版社, 2017.
- 17 Mayo Clinic. Melanoma - Symptoms and Causes [EB/OL]. [2021-03-10]. <https://www.mayoclinic.org/diseases-conditions/melanoma/symptoms-causes/syc-20374884>.
- 18 百度百科. 黑色素瘤 [EB/OL]. [2021-03-10]. <https://baike.baidu.com/item/%E9%BB%91%E8%89%B2%E7%B4%A0%E7%98%A4/507672?fr=aladdin>.
- 19 袁凯琦, 邓扬, 陈道源, 等. 医学知识图谱构建技术与研究进展 [J]. 计算机应用研究, 2018, 35 (7): 1929-1936.
- 20 The National Center for Biomedical Ontology. Prostate Cancer Ontology [EB/OL]. [2021-05-28]. <https://bioportal.bioontology.org/ontologies/PCAO/?p=summary>.
- 21 The National Center for Biomedical Ontology. Thyroid Cancer Ontology [EB/OL]. [2021-05-28]. <https://bioportal.bioontology.org/ontologies/TCO/?p=summary>.

(上接第 36 页)

- 2 Fraust A, Olive F. Preface of Virtual Special Issue on Smart Pattern Recognition for Medical Informatics [J]. Pattern Recognition Letters, 2019 (128): 146-147.
- 3 Jacob D, Chang M W, Lee K, et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding [J]. NAACL-HLT, 2019 (1): 4171-4186.
- 4 Cui Y, Che W, Liu T, et al. Pre-Training with Whole Word Masking for Chinese BERT [EB/OL]. [2019-06-30]. https://www.researchgate.net/publication/333892280_Pre-Training_with_Whole_Word_Masking_for_Chinese_BERT.

- 5 Liu Y, Ott M, Goyal N, et al. A Robustly Optimized BERT Pretraining Approach [EB/OL]. [2019-07-30]. https://www.researchgate.net/publication/334735779_RoBERTa_A_Robustly_Optimized_BERT_Pretraining_Approach.
- 6 Guo H H, Na X, Li J. Qcorp: An Annotated Classification Corpus of Chinese Health Questions [J]. BMC Medical Informatics and Decision Making, 2018, 18 (1): 16.
- 7 Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine Learning in Python [J]. JMLR, 2011 (12): 2825-2830.