

跨领域知识服务应用设计与实践探索^{*}

侯 丽 刘 燕 陈松景 胡志民

(中国医学科学院医学信息研究所 北京 100020)

〔摘要〕 从跨领域知识服务应用的内涵、特点、设计思路和实现路径等角度论述跨领域知识服务应用设计,以环境健康领域为例,介绍跨领域知识服务系统功能设计与开发,结合应用效果总结存在的问题并提出相应建议。

〔关键词〕 知识服务;跨领域应用;数据关联分析;数据整合

〔中图分类号〕 R-056 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2018.01.016

Design and Practical Exploration of Interdisciplinary Knowledge Service Application HOU Li, LIU Yan, CHEN Song-jing, HU Zhi-min, Institute of Medical Information of Chinese Academy of Medical Sciences, Beijing 100020, China

〔Abstract〕 From the perspectives like connotation, characteristics, design idea and realization path, the paper expounds on design of interdisciplinary knowledge service application; taking environmental health field as an example, it introduces the design and development of interdisciplinary knowledge service system function, summarizes the existing problems based on the application effect and puts forward corresponding suggestions.

〔Keywords〕 Knowledge service; Interdisciplinary application; Data association analysis; Data integration

1 引言

随着人工智能、智能手机、大数据技术等全新信息技术的涌现,新知识、新技术、新概念层出不穷,对数据以及信息资源的开发利用变得更加重要并逐渐发展成为一个产业。以数据和知识服务为主要支撑和动力的知识经济时代到来,信息服务已经由传统的以资料搜集与提供转变为对数据、各类资源的整合与深度服务,从以揭示单纯的信息为主要

服务方式,向以人、机、网相结合的以智力开发为主的数字化知识服务转化^[1]。数据密集型科学研究范式的出现更是对当下的知识服务模式提出更多新的挑战与机遇,不同领域的资源、数据以某种关系联结起来,挖掘出其深层次的关联;在新时代背景下提升知识服务的性能,保证用户的黏着性是值得信息工作者深入思考的问题。

中国工程科技知识中心是国家工程科技领域首个以跨领域专业数据融合与深度知识挖掘为特色的公益性、开放式资源集成和知识服务平台。自 2012 年启动建设以来,知识中心联合相关行业主管部门情报所及行业信息中心、国内顶尖高校、国家级科研院所、国内信息技术领先企业等构建了 30 多个专业领域知识服务系统,推动我国工程科技资源汇

〔修回日期〕 2017-12-12

〔作者简介〕 侯丽,副研究员,发表论文 20 余篇;通讯作者:胡志民,副研究员,发表论文多篇。

〔基金项目〕 中国工程科技知识中心建设项目“医药卫生专业知识服务系统”(项目编号:2017-2-1)。

聚、数据连通与知识融合,实现资源的高效配置和综合利用。为实现不同领域资源的汇聚、数据连通与知识融合,链接多方资源与数据共享的工具必不可少。跨领域知识服务应用在这样的背景下被提出,其应用场景明确,解决既有的众多科学技术领域平台中资源的关联、数据打通与共享,进而为更多工程科教研究人员提供深层次、专业的、集成的专指化服务。

近年来国内外大量学者对知识服务进行了相关研究,研究重点集中于以下两点:一是对知识服务定义的探讨。张晓林教授指出^[2]:知识服务是以信息知识的搜寻、组织、分析、重组的知识和能力为基础,根据用户的问题和环境,融入用户解决问题的过程中,提出能够有效支持知识应用和知识创新的服务。张红丽^[3]等指出:建立在相关服务人员的知识基础之上;提供给用户的可以是信息、知识或知识产品,甚至可以是解决方案,以解决用户具体而实际的问题为目标;追求知识服务对问题解决的价值效益,是知识服务机构的一种价值取向。总体而言,对知识服务定义可归纳为:以用户目标驱动,以解决用户需求为宗旨,以数据资源集成为核心的知识内容服务。二是对知识服务模式的探讨,大多集中对图书馆、信息服务机构的知识服务模式进行综合、分类探讨^[4-7],研究的侧重点在于培养具备专业素养、多重角色的馆员针对企业、学科、政府等用户提供不同的服务,其服务的基础依然是图书馆固有的传统资源及数据库。张晓林教授^[8]指出:知识服务的主要服务模式有专职顾问(1对1)服务模式、参考咨询服务模式和自助服务模式。随着数据库、知识库、大数据和云计算等相关技术的应用,构建具有自我知识检索、知识推理与分析的自助式知识服务平台已成为必然。对于跨领域服务的研究较少,仅有的与跨领域相关的研究大多集中在对跨界搜索以及图书馆的跨界服务上。陈昊天^[9]从搜索引擎技术的角度提出通过用户参与知识跨界搜索、企业伙伴参与跨界与产学研合作知识跨界搜索等方式来提升企业知识的应用能力及转化能力。张美莉^[10]从理论层面探索图书馆跨界服务的现状,提出拓展图书资源,充分利用互联网技术构建虚实

空间等线上线下相结合的知识平台。

综上,国内学者围绕传统图书馆的知识服务研讨较为完善,而结合不同领域的资源开展联合的“跨领域知识服务”开发与应用研究较少,而聚焦到以实现资源汇聚、数据连通与知识融合的“跨领域知识服务应用”相关研究更为薄弱。因此从实践层面而言,研究跨领域知识服务应用,能为打通国家工程科教领域覆盖农业、林业、医药、营养、环境等不同学科之间的信息壁垒提供实战场景。从发展前景而言,真正实现资源的跨界融合和知识创新服务,充分发挥知识中心在政产学研用全链条协同创新、集成创新、应用创新、服务创新的作用提供参考和借鉴。

2 跨领域知识服务应用内涵分析

2.1 概念界定

跨领域知识服务即是“跨领域”与“知识服务”的组合。结合中国工程科技知识中心现有分中心的领域特性,以及服务对象、平台既有功能的情况,开展跨领域知识服务。首先需要集合来自不同专业领域的机构,其次需要对其各自拥有的资源寻找连接点,最后基于连接点开展深度挖掘。因此,本研究认为跨领域知识服务应用是根据用户提出的非单一领域的知识服务需求,不同资源提供方结合自身领域优势和特点与其他资源提供方开展合作,利用两个或以上的领域资源,实现不同领域之间数据的贯通融合,揭示出单一领域数据源无法呈现出的深层次知识关联,为用户提供跨领域分析的创新型知识服务的工具和载体。

2.2 特点

依据以上概念界定,跨领域知识服务应用具有以下特点:(1)跨界性。跨领域知识服务应用一般由两个或以上资源建设方共同合作提供,利用多个领域资源,而不能仅局限于某一领域资源;所提供的知识服务应是跨学科、跨领域、跨行业的,着重体现出交叉跨界。(2)融合性。打通不同领域资源的数据源,实现相互关联,揭示内在联系;实现不

同领域、不同行业数据的汇聚、贯通与融合。(3) 协同性。涉及到的不同资源提供方之间要明确目标、任务与分工,充分沟通,密切合作,协作开发,共同开展服务。(4) 创新性。通过跨领域知识服务应用,揭示不同学科、领域、行业内关联并进行深入挖掘分析,提升知识服务水平,体现集成创新、应用创新、服务创新。

3 跨领域知识服务应用设计思路

3.1 设计原则

3.1.1 以用户需求为导向 跨领域知识服务应用的设计与开发必须以用户需求为导向,通过多种途径和方式开展用户需求调研,在明确用户实际需求的基础上有针对性地开展相关工作,为用户提供相关知识服务。

3.1.2 以数据资源为基础 数据资源是开展知识服务的基础和前提,服务方围绕目标通过既有数据资源整理或组织新建等方式,获取跨领域知识服务应用所需相应数据资源,同时资源建设方应保证数据资源能够持续保存和动态更新。

3.1.3 以关联融合为核心 不同领域的的数据资源经过汇聚打通,建立某种关联融合,形成纽带,才能实现不同领域间的数据联动和关系揭示,同时使合作各方间建立紧密联系,成为提供知识服务的有机整体。

3.1.4 以创新应用为目的 跨领域知识服务应用可以打通不同领域、行业数据资源,揭示出单一领域数据源无法呈现出的深层次知识关联并进行深入挖掘分析,创造更有意义、更有价值的知识服务,更好地满足用户对新知识的需求,更好地服务于工程科技创新发展。

3.1.5 协同合作,共建共享 跨领域知识服务应用调用不同领域的的数据资源,需要对不同领域的异构资源进行统一处理,包括数据格式转化、数据整合、数据存储等,因此相关资源建设和提供方必须共同设计、联合开发。跨领域知识服务应用的数据资源来自于不同资源提供方,但从用户的角度来看务必是一个完整的应用,因此对外必须提供统一的

服务接口,各资源提供方需紧密配合、共同服务。同时跨领域知识服务应用是不同资源提供方共同开发完成,共同享有合作成果,知识产权也应共同所有。

3.2 开发思路

结合文献调研及相关领域实践,本研究梳理了跨领域知识服务应用的开发流程,见图 1。

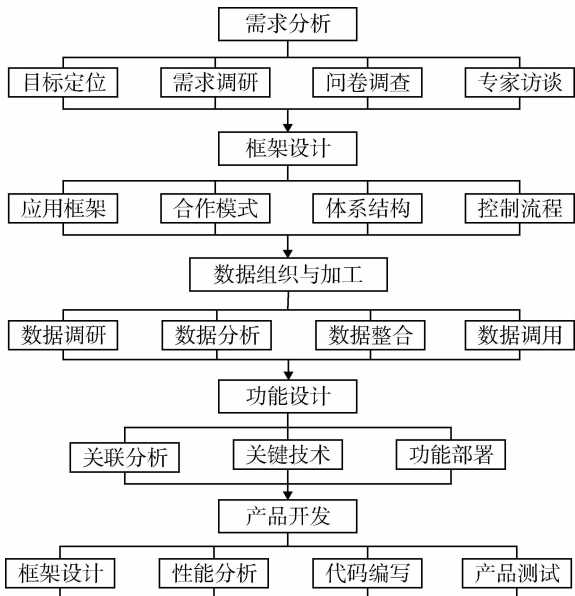


图 1 跨领域知识服务应用开发流程

完整的跨领域知识服务应用开发流程应包含 4 个环节:(1) 需求分析。通过实地调研、专家访谈、问卷调查等方式开展需求调研,结合跨领域数据资源现状,以用户需求及功能定位为导向,明确需求内容和建设目标。跨领域知识服务应用的需求获取方式、来源和目的都各不相同,例如根据课题组及科研团队的实际需求设计、开发跨领域知识应用,或根据现有资源提炼可提供的服务,或根据院士的要求与建议进行应用开发等。(2) 数据组织与加工。其包括数据采集、数据描述、数据分析(结构、内容等)、不同领域数据的整合、数据入库、数据管理、数据更新等。根据需求内容对采集到的数据进行清洗与过滤,制定数据存储标准,完成数据存储与入库。结合需求和已有数据资源,展开交叉领域数据的知识关联分析。(3) 功能设计。结合不同领域的的数据资源和特点,充分利用各领域的优

势,联合设计,为用户提供跨领域的数据关联服务,揭示数据深层的知识关联。(4) 产品开发。根据用户需求设计应用功能,完成应用的前端页面设计和后端功能开发并在此基础上搭建服务,供用户通过互联网进行访问和使用。设计并实现友好的数据展示方式,并对其进行功能测试,根据测试结果进行完善。

4 跨领域知识服务应用的实践

4.1 概述

环境健康是医学领域研究发展的一个重点交叉领域,环境对健康产生的影响越来越引起国家重视,积极持续开展环境健康研究也是实现“健康中国 2030”的重要保障。气象环境对人体健康有较大影响,尤其是空气污染物给人们健康带来较大危害,国际大量科学研究开始探索环境因素与疾病高发的关联关系^[11-13],国内环境健康相关研究大幅增长也证明了学界对该方面的关注^[14-16]。选取医药领域的疾病数据与地理、环境领域的地理信息、环境数据开展跨领域知识服务应用实践具有重要的现实意义。

4.2 数据采集

共采集资源约 27 万条,数据资源类型包含两大类。一类数据资源为疾病数据,包括国内呼吸科疾病数据,来自北京三甲医院急诊科就诊数据;慢性病数据(包括糖尿病、高血压等),来自国家卫生计生委每 5 年发布的国家卫生服务调查数据,包括疾病类型、年龄、性别、区域信息;国外疾病数据来自美国疾病控制与预防中心网站的开放科学数据^[17],包括美国各州肺癌发病率、白种人肺癌发病率、黑种人肺癌发病率等数据。另一类数据资源为环境气象数据,国家环保总局发布的空气污染物监测数据,包括空气污染物气体、细颗粒物、二氧化硫与碳等,以及气象数据,包括气温、湿度、气压等。国外环境数据来自美国国家环境保护署网站发布的 1990-2016 年美国环境数据,分别包括各州按年度、月份、天记录的一般污染物气体(如臭氧 O-

zone、二氧化硫 SO₂、一氧化碳 CO、二氧化氮 NO₂)、细颗粒物(如 PM2.5、PM10)、气象指标(风速、温度、大气压等)、有毒物质(有毒空气污染物、挥发性有机污染物、氮氧化合物等)。

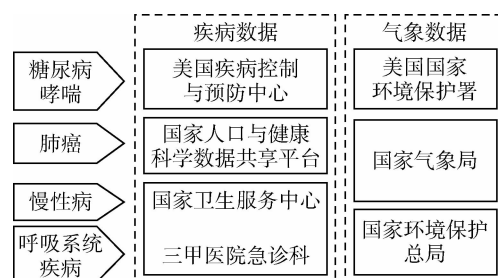


图2 环境健康跨领域数据分布

4.3 数据整合与关联分析

如上科学数据来自不同机构、不同领域,其结构差异大、数据类型多样。结合数据特征制定数据处理流程如下:(1) 数据预处理。对采集到不同领域的数据分别进行清洗、去噪、降维、空缺值处理、过滤等。(2) 数据存储标准制定。结合健康、环境两个跨领域数据的数据属性及资源特征,制定跨领域数据存储标准,完成数据存储与入库。(3) 数据资源整合。选取健康与环境数据之间的关联关系及相关规则,制定跨领域资源整合方案。依据特定规则对来自不同领域的的数据资源进行整合,实现跨领域数据资源整合。(4) 数据关联分析。利用美国呼吸系统疾病数据、慢性病数据和环境气象科学数据等资源,展开呼吸系统疾病与气象因素、呼吸系统疾病与空气污染指标、慢性病发病区域特征与肿瘤发病区域特征等方面的交叉领域知识关联分析。(5) 关联方法选择。本研究对整合数据的分析方法采用关联分析方法构建危险因素与疾病发病率的关联关系,进而为可视化分析提供数据基础。关联关系挖掘采用 Apriori 算法,经过逐层搜索迭代方法,挖掘出污染物气体、细颗粒物、气象因素、有毒物质等环境因素与肺癌、哮喘、糖尿病等关联关系,开展关联分析^[18]。

4.4 基于数据分析的环境健康服务模块功能设计

为实现采集数据的综合应用与服务,本研究设

计了可用于人机交互展示的系统功能架构,并对前文描述的各种数据整合与关联关系进行平台系统实现,在“医药卫生知识服务系统(<http://med.ckcest.cn>)”中新增“环境健康”专题服务,该专题知识服务系统架构,见图 3。环境健康跨领域知识服务应用开发主要针对呼吸系统疾病、慢性病、肿瘤等疾病,提供各类型疾病发病地理分布分析功能;提供疾病与气象因素、空气污染指标等多维度关联分析功能;提供时间维度的疾病发病变化分析功能,包括纵向的不同年份之间疾病发生对比分析功能、横向的同年份不同区域疾病发生对比分析功能;提供不同疾病类型的国内外发病情况对比分析功能。开发的平台功能界面,见图 4。

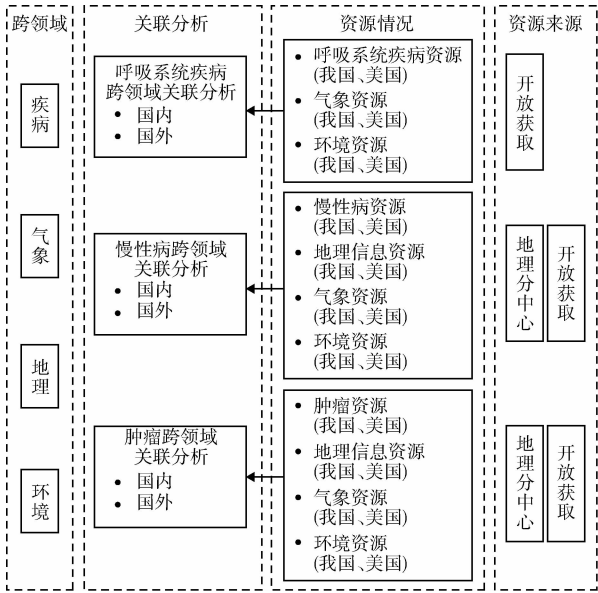


图 3 跨领域知识服务总体框架

4.5 跨领域知识服务应用效果分析

该跨领域知识服务应用实现了气象因素与呼吸系统疾病关联分析、空气污染物与呼吸系统疾病关联分析、呼吸系统疾病发病人口学特征分析(国内部分),以及美国加州哮喘与空气污染物关联分析、美国各地区哮喘与气象环境关联关系、美国各州气象环境数据可视化分析(国外部分)。揭示了国内

外呼吸系统疾病发病与对应地区气象环境间的关联关系,实现了糖尿病患病率区域特征分析、高血压患病率区域特征分析(国内部分),以及美国各州糖尿病患病率统计分析、美国各州糖尿病患病率与气象环境关联分析等(国外部分)。揭示了国内外常见慢性病与对应区域气象环境间的关联关系,实现了肿瘤发病区域特征分析、肺癌发病区域特征分析等(国内部分),以及美国各州肺癌发病率统计分析、美国各州肺癌发病率与气象环境关联分析等(国外部分)。揭示了国内外肿瘤与对应区域气象环境间的关联关系。围绕常见高发疾病的区域特征分析,以及针对特定疾病高发区域开展疾病与气象环境的关联分析,能为用户提供直观的数据展示,帮助用户直观了解各类疾病影响因素。本研究以空气污染物对疾病发病率的关联为例进行说明:进入“空气污染物与呼吸系统疾病关联分析”,选取“发病人数随 PM2.5 浓度变化”,可发现随着 PM2.5 浓度的增加,急性上呼吸道感染、感冒和肺炎等疾病的发病率呈显著上升趋势。急性下呼吸道感染类疾病与 PM2.5 浓度变化不存在特别显著的正相关影响。PM2.5 与呼吸道疾病感染关联关系,见图 5。由此可见,PM2.5 对上呼吸道疾病感染及感冒肺炎等常见呼吸系统疾病有显著正相关影响。通过该平台的后台日志分析,发现“环境健康”知识服务应用模块位居“医药卫生知识服务系统”访问量的最高,由此说明跨领域知识服务应用符合广大工程科教领域人员的关注点,能解决用户对医药数据中潜在关联关系的发现。结合不同领域的数据开展知识关联与挖掘,对科研人员了解疾病发病、多因素关联,开展创新研究提供新的视角并在一定程度上通过对分析数据的解读,帮助公众提高对环境健康的认知,为减少或预防高发区域内某种疾病的发生提供依据,为国家工程科技智库在医药卫生、公众健康、科技创新和产业发展等方面开展宏观发展策略研究提供全面、开放、智能的知识资源和知识服务,为我国医药卫生事业发展做好服务支撑工作。



图4 跨领域知识服务应用系统页面

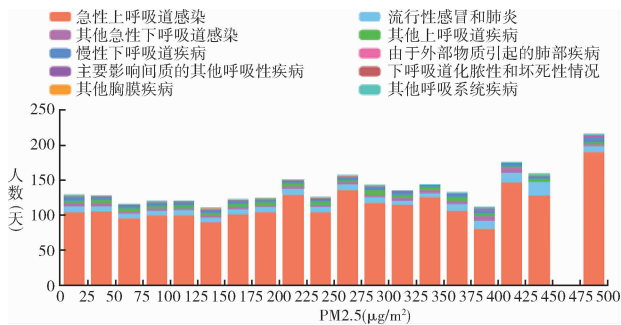


图5 PM2.5与呼吸道疾病感染关联

5 结语

环境健康跨领域知识服务应用整合国内外环境与健康跨领域资源，涉及呼吸系统疾病、哮喘、糖尿病、高血压、肺癌等多病种，气象、环境、地理信息等多领域，结合年龄划分、性别、疾病发病率和死亡率等多维度，开展国内外多个地区气象条件、空气污染情况与疾病情况的关联知识服务，开拓国内相关知识服务机构基于多病种多维度关联的深度知识挖掘应用。同时与地理信息分中心、环境分中心开展跨领域合作，基于已有资源针对特定区域进行疾病与环境关联分析，实现对比不同区域的跨领域多病种多维度的关联知识服务，并且提供国

内外呼吸系统疾病、肿瘤、慢性病等多病种发病、死亡情况与气象、环境、地理等关联关系对比，为宏观决策和科学研究提供了有利的支撑依据和重要参考。

作为一个初步尝试的知识服务应用，环境健康知识服务应用取得了一定效果，但是若要将这种服务模式在中国工程科技知识中心平台，甚至在其他行业推广，还需要在多个层面加强合作与实施。可通过如下渠道提升跨领域资源建设质量：（1）合作单位协作建设数据资源，围绕建设目标共同采集和整理数据，或根据实际需求重新对已有数据进行组织加工，保证数据资源间的相互匹配度与关联性。（2）加强数据资源全生命周期管理，保证数据资源的连续性和完整性。（3）进一步推进合作单位间的数据资源合作与共享机制建设，扩大开放共享使用。随着中国工程科技知识中心工作的不断深入，各领域资源汇聚成“海”，打通“数据海”。建立知识关联，形成跨领域的知识融合，提供创新型知识服务，是知识中心发展的重要方向，也是知识中心不同于其他知识服务平台的关键所在、优势所在。跨领域知识服务应用正是实现知识中心价值的重要工具，在后续工作中还将不断深入探讨和完善，使其发挥更大作用。

参考文献

- 1 代涛, 钱庆, 王小万, 等. 医疗卫生领域知识服务与知识管理的理论和实践 [J]. 医学信息学杂志, 2008, 29 (4): 1-10.
- 2 张晓林. 走向知识服务 [M]. 成都: 四川大学出版社, 2001: 110-112.
- 3 张红丽, 吴新年. 知识服务及其特征分析 [J]. 图书情报工作, 2010, 54 (3): 23-27.
- 4 刘佳. 高校图书馆学科知识服务模式研究 [D]. 长春: 吉林大学, 2007.
- 5 董玮, 詹庆东. 图书馆知识服务模式辨析 [J]. 图书馆学研究, 2016, (3): 72-79.
- 6 张海涛, 宋拓, 刘健. 高校图书馆一站式知识服务模式研究 [J]. 情报科学, 2014, 32 (6): 104-108, 113.
- 7 马天舒. 大学图书馆开放型知识服务模式研究 [J]. 情报杂志, 2013, 32 (2): 135-138, 102.
- 8 张晓林. 走向知识服务: 寻找新世纪图书情报工作的生长点 [J]. 中国图书馆学报, 2000, 26 (129): 32-37.
- 9 陈昊天. 跨界搜索对服务创新绩效的影响研究 [D]. 杭州: 浙江理工大学, 2016.
- 10 张美莉. 基于“互联网+”思维的图书馆跨界服务研究 [J]. 情报杂志, 2016, 35 (9): 204-207, 198.

- 11 Atkinson R W, Kang S, Anderson H R, et al. Epidemiological Time Series Studies of PM2.5 and Daily Mortality and Hospital Admissions: a systematic review and meta-analysis [J]. Thorax, 2014, 69 (7): 660-665.
- 12 Franchini M, Guida A, Tufano A, et al. Air Pollution, Vascular Disease and Thrombosis: Linking Clinical Data and Pathogenic Mechanisms [J]. Journal of Thrombosis and Haemostasis, 2012, 10 (12): 2438-2451.
- 13 Shang Y, Sun Z, Cao J, et al. Systematic Review of Chinese Studies of Short-term Exposure to Air Pollution and Daily Mortality [J]. Environment International, 2013, 54 (4): 100-111.
- 14 屈芳. 环境气象因素对呼吸系统疾病影响的研究进展 [J]. 气象科技进展, 2013, 3 (6): 35-44.
- 15 赵桂柱, 张冬梅. 论肿瘤与环境的关系 [J]. 中国中医药现代远程教育, 2012, 10 (12): 91-92.
- 16 郭堂春. 华中区域常见慢性非传染性疾病前瞻性队列研究 [J]. 中华预防医学杂志, 2016, 50 (10): 927.
- 17 美国疾病控制与预防中心 [EB/OL]. [2017-04-20]. <https://www.cdc.gov/>.
- 18 陈松景, 吴思竹, 侯丽. 基于疾病与环境科学数据的跨领域关联分析及应用 [J]. 中华医学图书情报杂志, 2017, 26 (7): 1-6.

(上接第 54 页)

态的健康建议方案, 推广“专科特征+平台互动”的健康互联网应用, 推动“互联网+医疗”在妇幼保健领域的应用, 为智能问诊软件系统开发和孕产智能陪护机器人实体研发奠定核心基础。由于孕产智能问诊知识库是从科研成果转化到实践应用, 需要一个逐步完善与应用推广的过程, 需经历较长时间的不断知识积累与实践验证才能形成可靠有效的应用。

参考文献

- 1 丁腊春, 王译, 石雷, 等. 面向孕产妇的移动医疗健康平台“孕乐宝”开发与应用 [J]. 医学信息学杂志, 2016, 37 (10): 40-43.
- 2 吴佳男. 智能医疗: 或成医疗机构新突破口 [J]. 中国医院院长, 2017, (2): 89.

- 3 常晓. AI+医疗 [J]. 人人健康, 2016, (21): 22-23.
- 4 瞿裕忠, 胡伟, 程龚. 语义网技术体系 [M]. 北京: 科学出版社, 2015: 10.
- 5 邱旭. 领域本体在糖尿病领域知识服务中的应用 [D]. 哈尔滨: 哈尔滨工业大学, 2013.
- 6 李博. 基于语义关系的高血压临床指南知识库构建 [D]. 成都: 电子科技大学, 2013.
- 7 陈华钧, 姜晓红, 吴朝晖. 支持中医药信息化的语义网格平台实现 [M]. 杭州: 浙江大学出版社, 2011: 6.
- 8 吴朝晖, 陈华钧. 语义网格: 模型、方法与应用 [M]. 杭州: 浙江大学出版社, 2008: 5.
- 9 范启勇. 上海市医院信息集成平台建设与实践应用指南 [M]. 北京: 科学出版社, 2016: 11.
- 10 来建梅. 基于领域本体的中医骨伤知识库的研究与构建 [D]. 济南: 山东中医药大学, 2013.