

# 新型冠状病毒肺炎防控专题服务构建研究\*

刘燕 李姣 康宏宇 李露琪 张潇潇 徐子隽 侯丽

(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]** 详细阐述基于资源聚合与可视化分析技术的新型冠状病毒肺炎防控专题构建方案, 包括需求分析、模块设计、架构搭建、数据采集与处理、可视化分析、功能开发与实现等流程, 为社会公众、科研人员、领域决策者等提供权威、全面、及时的新冠肺炎综合知识服务。

**[关键词]** 新型冠状病毒肺炎; 专题服务; 资源聚合; 可视化分析

**[中图分类号]** R-056 **[文献标识码]** A **[DOI]** 10.3969/j.issn.1673-6036.2020.08.003

**Study on the Building of the COVID-19 Epidemic Prevention and Control Thematic Service** LIU Yan, LI Jiao, KANG Hongyu, LI Luqi, ZHANG Xiaoxiao, XU Zidu, HOU Li, Institute of Medical Information, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100020, China

**[Abstract]** The paper elaborates the building scheme of the COVID-19 epidemic prevention and control thematic service in detail based on resource integration and visualization analysis technology, including the processes of requirements analysis, module design, architecture establishment, data collection and processing, visualization analysis, function development and implementation, etc., so as to provide the COVID-19 knowledge services authoritatively, comprehensively and timely for the public, researchers, decision-makers in the field and other personnel.

**[Keywords]** COVID-19; thematic service; resources integration; visualization analysis

## 1 引言

### 1.1 研究背景

2019 年 12 月湖北省武汉市暴发新型冠状病毒肺炎(以下简称新冠肺炎)疫情, 国家卫生健康委员会(以下简称国家卫健委)等有关部门紧急采取

一系列新冠肺炎防控<sup>[1]</sup>、发布诊疗方案<sup>[2]</sup>等措施, 全国多个省份启动一级响应。与此同时世界卫生组织(World Health Organization, WHO)不断上调全球疫情风险级别并称新冠肺炎可被定义为全球性大流行病<sup>[3]</sup>。与日俱增的新冠肺炎确诊病例数, 为一线医护人员、疫苗、药物研发科研团队、相关决策层带来巨大挑战, 给广大公众生活和工作带来极大不便与一定程度的恐慌。同时网络上关于疫情的信息很多, 公众一时很难接受并分辨其准确性与权威性。此外有研究指出科研人员用于收集资料的时间占整个科研过程的 51%, 有效的时间和精力不能被充分用于关键研究中发现<sup>[4]</sup>。因此亟需构建一个整合式的新冠肺炎防控专题服务平台, 实现一站式疫情动态、疫情数据分析、科普知识、权威报道、权

**[收稿日期]** 2020-06-26

**[作者简介]** 刘燕, 硕士, 助理研究员, 发表论文 3 篇; 通讯作者: 侯丽, 博士, 副研究员。

**[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: CKCEST-2020-1-14)。

威医学文献、调研报告、指南规范、疫情研究、临床资料、公众关注热度、确诊患者同行程查询工具等各方面实时数据和权威知识的集中发现与获取服务,方便各个层面用户通过一站式整合知识服务平台各取所需,进而为疫情防控与研究提供全面、多元的资源整合和知识服务。

## 1.2 研究团队

中国工程院中国工程科技知识中心医药卫生专业知识服务系统作为医药知识服务专业平台,在异构资源整合、关联、分析、挖掘等方面具备丰富技术积累。为响应习近平总书记对新冠肺炎疫情的重要指示,以及国家卫健委对信息准确发布、客观报道疫情进展、快速传递新冠肺炎相关知识的要求,充分发挥团队在医药卫生领域信息支撑与知识服务的作用,本研究基于已有技术积累与建设经验及时开展疫情资讯、防控指南、政策、文献等的集成与相关数据可视化分析,快速研究并构建新冠肺炎防控专题服务,进而为公众、医药科教人员、决策者等提供权威、专业、全面、多元的疫情信息与知识获取服务。

## 2 相关研究

### 2.1 国内已有研究

自新冠肺炎暴发以来,国内外学者陆续对其临床表现、实验室特征、传播趋势、治疗方案、新药研发、防控防治等方面进行深入研究,相关研究成果不断推出。为方便研究人员快速掌握科研进展,多家机构从不同角度搭建一系列专题和平台,重点围绕新冠肺炎相关科技文献、科研动态等资源提供服务。国家科技图书文献中心重点领域信息门户搭建新冠病毒动态监测平台<sup>[5]</sup>,及时整理、展示最新的相关科技资源。万方医学针对国内外疫情相关文献进行研究整理,开发新冠肺炎专题频道<sup>[6]</sup>,持续提供最新、权威的国内外疫情学术资料。中国科学院文献情报中心整合发布新型冠状病毒科研动态监测平台<sup>[7]</sup>,重点整理病毒方面的最新科研进展。此外为方便医护人员、一线工作者等开展工作,降低交叉感染风险,相关平台及信息系统也不断涌现。

中国电子技术标准化研究院开发并上线面向新冠肺炎的防疫用品标准化信息服务平台<sup>[8]</sup>,为医用防护用品标准规范知识查询提供便利。山西省汾阳医院郭庆峰等基于“互联网+”信息技术构建新冠肺炎疫情防控人员管理平台<sup>[9]</sup>,实现对隔离病区医务人员全程信息化管理。西安交通大学口腔医院李艺凡等构建新冠肺炎流行期预检分诊信息系统<sup>[10]</sup>,有效缩短患者分诊时间,降低交叉感染风险。另外中国中医科学院中药研究所郭非非等设计并开发中药抗新冠肺炎药效预测分析平台<sup>[11]</sup>,以实现中药或方剂潜在药效作用的快速预测、分析。

### 2.2 存在的不足

综上所述,目前已开发并应用不少科研动态监测平台、应用服务平台等,但鲜有针对疫情资讯、统计数据、专业指南、科普知识、学术研究、政策法规、防控知识、临床资料、专题快报、相关工具等全方位知识进行整合与挖掘的综合性专题服务,无法满足不同层面用户对于一站式获取疫情发展动态与防护知识的需求。加之各种渠道传播的科普、防护等知识层出不穷,用户筛选权威、有价值信息的成本不断增高。因此面向社会大众、科技工作人员、决策管理者等开展全面、综合的知识服务专题建设迫在眉睫。本研究依托于中国工程科技知识中心医药卫生专业知识服务系统丰富的专题知识服务整合与构建经验,希望在面向新冠肺炎防控专题服务的研究与构建方面做些尝试,以期为公众获取疫情相关知识提供便利,为科研人员进一步开展深入挖掘与分析提供数据支撑,为决策管理者制定防控方案、工作部署等提供情报服务。

## 3 面向资源聚合的专题知识服务构建方案研究

### 3.1 需求分析

通过对用户需求的深入调研发现,面对突如其来的新冠肺炎疫情,公众、科技工作者、决策管理者等不同层面的用户对信息的需求不一致。其中公众最关心的是全国及自身所在省市的实时疫情动态,国家有关部门针对疫情发布的防控措施与政策

指南, 疫苗研究与投入临床应用的进展情况, 病毒相关科普知识, 自身应该如何进行防护, 同行程是否有确诊患者以及定点医疗机构与发热门诊信息等; 科技工作者则更关注科研领域的最新研究与发现, 如病原研究、药物与疫苗研发、临床诊疗进展等相关的前沿科技文献与科技报告; 决策管理者需要的是指南规范、防控措施、专题报告、重要报告等可以辅助其制定决策的知识。此外现有各网站资源类型都相对单一、分散, 用户很难在短时间内找到全方位的资源, 时间成本耗费普遍较高。鉴于以上需求, 从多源数据采集出发, 制定面向不同用户

的专栏, 设计面向不同数据的整合模型与方法, 以期最大程度满足不同层面用户对疫情知识的一站式获取需求, 构建面向资源聚合的新冠肺炎防控专题知识服务。

### 3.2 模块设计

3.2.1 概述 通过需求分析、文献调研、专家咨询等方法初步确定新冠肺炎防控专题知识服务页面设置方案, 包括9大知识版块, 见图1。用户可通过快速浏览来准确定位其感兴趣的内容以便深入学习, 从而有效降低资源检索与发现的时间成本。

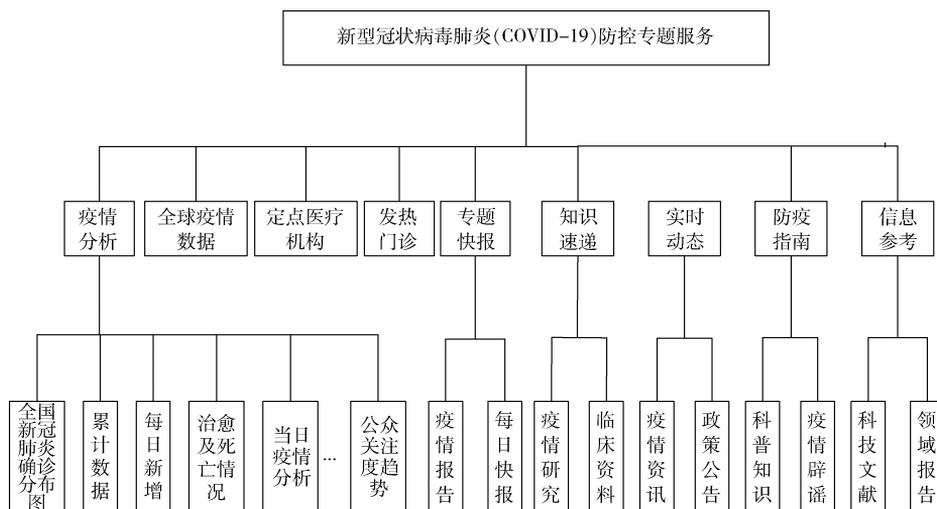


图1 新冠肺炎防控专题服务页面设置方案

3.2.2 疫情分析 定期采集并更新国家卫健委、中国疾病预防控制中心等官方网站发布的每日疫情数据, 包括全国各省、湖北各市、北京各区的确诊、疑似、死亡、治愈病例数等, 以及百度等搜索引擎统计的公众关注度数据, 利用数据挖掘技术和可视化工具进行多维度数据分析与展示, 方便用户及时、便捷地掌握疫情动态, 多角度直观了解疫情发展趋势。

3.2.3 全球疫情数据分析 每日采集并更新世界卫生组织发布的疫情数据, 包括各国累计确诊病例数、死亡病例数、死亡率等, 以地图形式直观展示, 对相关数据按指标进行排序, 便于用户查看全球疫情发展态势。

3.2.4 定点医疗机构 集成全国33个省、市、自治区卫健委、卫健局等官方机构公开发布的新冠肺

炎定点医疗机构数据, 包括所在省份、城市、辖区、医疗机构名称、地址等信息, 为新冠肺炎定点救治医疗机构浏览、选择、查询等提供帮助。

3.2.5 发热门诊 整合全国33个省、市、自治区卫健委、卫健局等官方机构公开发布的新冠肺炎发热门诊数据, 包括所在省份、城市、辖区、医疗机构名称、地址等信息, 为新冠肺炎发热门诊的浏览、选择、查询等提供便利。

3.2.6 专题快报 整合由专业队伍定期搜集整理的新冠肺炎疫情相关分析报告和面向中国工程院院士整理的每日疫情专题快报, 包括国内外疫情变化趋势、快报概述、新闻资讯、专家观点、科研动态等, 旨在为用户提供精品知识和深度分析内容。

3.2.7 知识速递 收集、整理新冠肺炎相关病原病因、药物疫苗等最新研究成果及临床诊疗研究资

料,方便有关人员及时掌握同行研究进展,提供更专业的疫情知识速递服务。

3.2.8 实时动态 筛选、整合世界卫生组织、国家卫健委等国内外权威网站发布的疫情实时资讯和国家相关政策公告,为用户第一时间了解疫情发展与防控防治情况提供窗口。

3.2.9 防疫指南 汇聚科普中国、健康中国等发布的权威新冠肺炎防护科普图文以及疫情辟谣知识,提供专业、可靠的健康防护知识和专家解读,以帮助用户正确、科学地防控疫情。

3.2.10 信息参考 收集、整合关于新冠肺炎国内外顶级医学期刊网站最新发表的研究成果、未经同行评议自愿提交到预印本平台的论文手稿以及领域报告,方便有关人员及时查阅科研进展资料,为开

展进一步研究提供参考。此外专题集成新冠肺炎确诊患者同行程、确诊患者活动小区、定点医院导航查询等便捷工具,帮助用户获得更全面的疫情信息。

### 3.3 总体架构

新冠肺炎防控专题服务总体架构包括数据采集层、数据处理层、专题模块层和专题交互层,见图2。采用人工与机器相结合的方式收集数据,利用大数据处理技术进行数据规范化处理与加工,借助数据挖掘与可视化智能分析工具开展不同维度的疫情数据分析,依据专题模块设计方案实现多源异构数据整合,完成专题页面框架搭建与功能开发,最终面向用户提供资源浏览、获取等功能以及数据可视化交互、工具查询应用等服务。

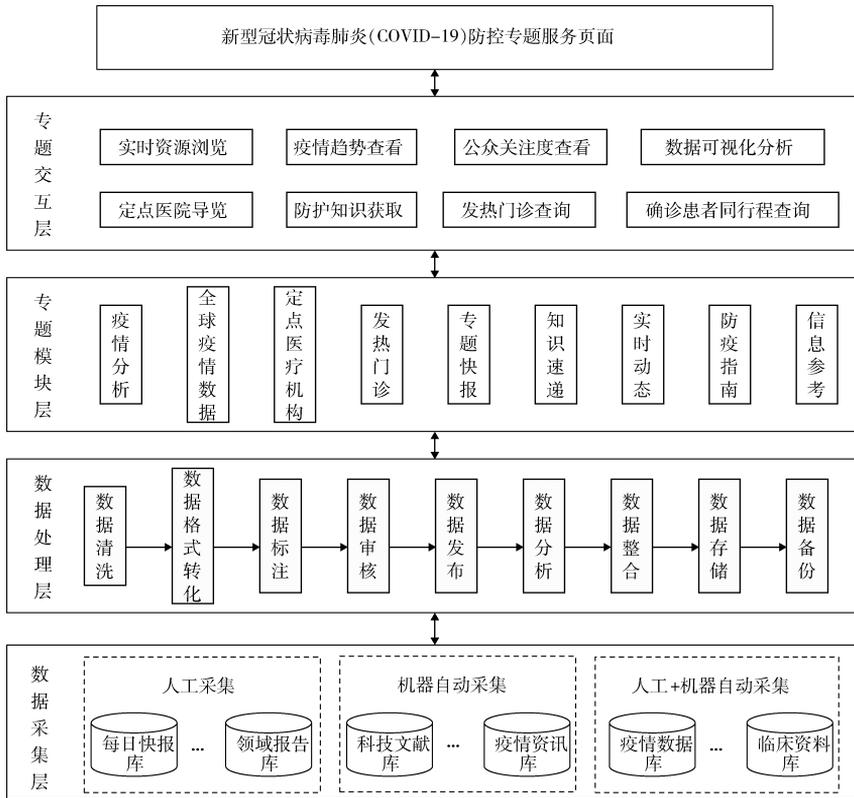


图2 新冠肺炎防控专题服务总体架构

## 4 专题服务构建关键技术

### 4.1 数据采集

本研究依据新冠肺炎各方面数据来源、自身结

构等特征以及是否可以满足专题展示与分析要求,将数据分为人工、机器自动、人工和机器采集相结合3种方式。人工采集方式包括采集来源调研与分析、采集字段整理、数据采集与下载等一系列流程,具有数据质量高、采集成本大等特点,主要针

对自建或自加工的资源,如每日快报、疫情报告、领域报告等;机器自动采集方式需经过采集来源梳理、模板制定、接口管理等一系列操作,具有采集成本低等特点,主要面向结构相对固定、数据量较大的资源,如科技文献、疫情资讯、科普知识等;人工和机器相结合的采集方式主要面向疫情实时数

据以及需要人工审核的数据类型,如疫情数据、疫情研究、临床资料等。专题涵盖的各类资源均来源于国内外多个权威网站,采集字段全面,采集频率高,能在一定程度满足各层面用户需求。截至2020年5月19日专题各类资源采集情况,见表1。

表1 专题各类数据采集情况

知识版块	资源类型	采集方式	来源	频率	语种	数据量
疫情分析	全国及各省市疫情数据	人工+自动	国家卫健委、中国疾病预防控制中心、湖北省卫健委	1次/天	中文	17 244 条
	公众关注指数	人工	腾讯、百度、新浪			1 071 条
全球疫情数据	-	人工+自动	世界卫生组织	1次/天	中文	26 190 条
定点医疗机构	-	人工	国家人口健康科学数据中心数据仓储	1次/周	中文	2 683 个
发热门诊	-	人工	国家人口健康科学数据中心数据仓储	1次/周	中文	11 765 个
专题快报	疫情报告	人工	自建	1次/周	中文	34 期
	每日快报	人工	自建	1次/天	中文	78 期
知识速递	疫情研究	人工+自动	《国际传染病杂志》、《中华疾病控制杂志》等	1次/天	中文/英文	46 条
	临床资料	人工+自动	《柳叶刀》、《疫苗杂志》等	1次/天	中文/英文	47 条
实时动态	疫情资讯	自动	新华网、人民网等	2次/天	中文/英文	956 条
	政策公告	自动	国家卫健委、疾病预防控制局等	1次/天	中文	126 条
防疫指南	科普知识	自动	国家卫健委、科普中国等	1次/天	中文	1 523 条
	疫情辟谣	自动	国家卫健委、宣传司、健康中国等	2次/周	中文	559 条
信息参考	科技文献	自动	medRxiv, bioRxiv,《柳叶刀》等	1次/天	中文/英文	10 497 篇
	领域报告	人工	新华网、人民网等	2次/周	中文/英文	88 份

## 4.2 数据处理

本研究利用大数据处理技术进行数据规范化处理与加工,首先通过医药卫生专业知识服务系统的数据加工平台完成噪音、冗余数据清洗,之后依据资源描述规范与标准制定统一数据规范,完成格式转化,制定数据标签并赋予唯一ID,最后采取分布式存储技术实现数据入库与备份,相关数据均存储于MySQL数据库中。此外对于部分人工采集、加工的数据可通过数据加工平台进行上传、审核、发布,进而实现多源异构数据整合。鉴于本研究涉及的数据类型和来源较多,因此要针对每类资源制定审核与入库规范,当有冲突时应及时进行处理与加工,具体包括不同来源数据字段格式不一致和数据重复两个方面。对于多来

源数据,普遍存在字段格式不一致的情况,如日期格式有多种表达形式,需根据展示方式进行同语种不同形式之间的转换、中英文之间的转换等规范化处理。对于多来源数据存在重复的情况,需根据制定好的规则进行重复值判定与筛选,如疫情资讯版块中多网站对同一事件的集中报道。此外针对系统自动采集的数据需进行空值判定与处理,若出现文献摘要等重要字段为空的数据时进行人工补充或删除不合格数据。对于文献作者等存在多值情况的字段,需基于人工与机器的方式进行冗余数据处理与正确性验证。

## 4.3 可视化分析

为方便相关人员研究疫情拐点、研判疫情趋势、洞察公众关注热度、挖掘各维度数据之间关联

关系,本研究通过对疫情数据的全面调研和整合分析,设计丰富、多维度智能分析模型,借助可视化分析工具对数据进行关联分析,以期为用户提供直观、生动、可交互的数据分析服务。疫情分析模型结合用户需求,从多个角度对确诊、死亡、疑似病例等热点数据进行对比分析,具体包括全国新冠肺炎确诊分布图、累计病例数据和每日新增病例分析、治愈和死亡趋势、全国疫情趋势对比、当日疫情分析、公众关注度趋势分析,以及全球疫情数据分析等多个维度。此外为突出重点地区疫情情况,对湖北省和北京市进行更为细致的分析。利用 ECharts 工具对疫情数据进行深入挖掘。该工具功能较为强大,涵盖地图、散点图、折线图、柱状图、环形图等多种图表,提供标题、图例、时间轴等可交换组件,支持多图表、组件的联动和混搭展示,赋予用户挖掘和整合数据的能力。

### 5 专题服务功能开发与实现

本研究以加工后的数据为基础,借助数据挖掘技术与可视化分析工具,使用 Java 编程语言进行专题服务功能开发与实现,于2020年2月7日正式上线对外提供服务。新冠肺炎防控专题服务页面(<http://med.ckcest.cn/covid-main.html>),见图3。



图3 新型冠状病毒肺炎防控专题服务页面

用户可进行相应资源浏览与获取。以地图方式呈现全国各地和世界各国累计确诊病例数,颜色深

浅代表该地或该国疫情严重程度,颜色越深说明越严重;以折线图方式展示全国、北京市随时间推移疫情的变化趋势,不同颜色代表不同维度疫情变化情况;以柱状图来对比分析北京各区、北京与其他省的疫情状况,不同颜色代表不同维度的数据;同时对公众关注度进行比较分析,见图4。

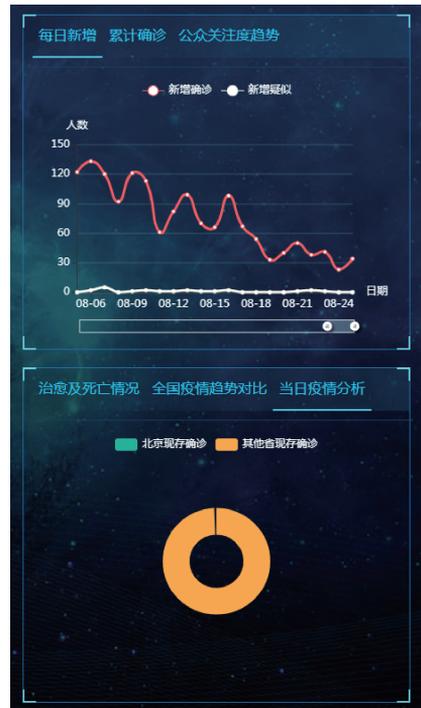


图4 疫情分析示例

为方便相关人员快速获取新冠肺炎定点医疗机构与发热门诊数据,专题按省份和地区进行细致的组织归类。此外为满足用户更精确的查找需求,对专题快报、知识速递、实时动态、防疫指南、信息参考等各类资源进行细致分类与设计,所有资源均按照时间倒序排列展示,可通过点击来源进行原始数据的查看与学习。同时专题尽可能提供数据全文,以方便用户下载。经过一段时间的运行与对外服务,资源覆盖面不断扩大,数量日益增长,功能更加完善,为用户一站式获取新冠肺炎相关知识提供渠道,得到用户一致好评与认可。

### 6 结语

本研究结合用户需求与专家建议,明确专题服

务目标和定位, 确定涵盖资源类型与相应数据采集来源, 设计页面框架及展示方式, 收集、整合各栏目权威数据资源, 挖掘、分析疫情数据, 快速开发并构建面向不同层面用户的新冠肺炎防控专题服务。有效缩短公众获取可靠来源信息的时间, 便于用户直观查看疫情发展态势以及公众关注度变化趋势, 为科技工作者发现高价值的知识服务奠定基础, 但还存在资源无法检索、未实现个性化推送、数据挖掘与利用不够充分等缺陷。专题将持续更新与完善, 积极发挥在医药卫生领域信息支撑与知识服务的作用。在此也呼吁社会各界充分发挥各自优势全力推动疫情防控工作, 争取早日打赢这场战“疫”。

## 参考文献

- 1 国家卫生健康委办公厅. 新型冠状病毒肺炎防控方案(第六版) [EB/OL]. [2020-05-19]. <http://www.nhc.gov.cn/xcs/zhengcwj/202003/4856d5b0458141fa9f376853224d41d7/files/4132bf035bc242478a6eaf157eb0d979.pdf>.
- 2 国家卫生健康委办公厅, 国家中医药管理局办公室. 新型冠状病毒肺炎诊疗方案(试行第七版) [EB/OL]. [2020-05-19]. <http://www.nhc.gov.cn/xcs/zhengcwj/202003/46c9294a7dfe4cef80dc7f5912eb1989/files/ce3e6945832a438eaae415350a8ce964.pdf>.
- 3 World Health Organization. WHO Director - General's Opening Remarks at the Media Briefing on COVID - 19 - 11 March 2020 [EB/OL]. [2020-05-19]. <https://www.who.int/dg/speeches/detail/who-director-general-opening-remarks-at-the-media-briefing-on-covid-19--11-march-2020>.

- 4 文庭孝, 刘晓英. 基于引文分析的我国研究者信息获取能力评价研究 [J]. 图书与情报, 2011 (6): 21-25, 32.
- 5 国家科技图书文献中心. 重点领域信息门户新发突发病(新型冠状病毒肺炎) [EB/OL]. [2020-05-19]. <http://portal.nstl.gov.cn/STMonitor/qbwnw/openhome.htm?serverId=26>.
- 6 北京万方数据股份有限公司. 新型冠状病毒肺炎专题频道 [EB/OL]. [2020-05-19]. <http://subject.med.wanfangdata.com.cn/Channel/7>.
- 7 中国科学院文献情报中心. COVID - 19 科研动态监测 [EB/OL]. [2020-05-19]. <http://stm.las.ac.cn/STMonitor/qbwnw/openhome.htm?serverId=172>.
- 8 《电子世界》杂志编辑部. 电子标准院上线新冠肺炎防疫防护用品标准信息服务平台 [J]. 电子世界, 2020 (4): 4.
- 9 郭庆峰, 杨扣琴, 赵丽婷, 等. 基于“互联网+”新型冠状病毒肺炎疫情防控人员管理平台的构建 [J]. 护理研究, 2020, 34 (5): 753-755.
- 10 李艺凡, 孙辉, 胡志萍, 等. 口腔专科医院新型冠状病毒肺炎流行期预检分诊信息系统的构建与应用 [EB/OL]. [2020-05-19]. <http://kns.cnki.net/kcms/detail/61.1399.R.20200224.0944.012.html>.
- 11 郭非非, 张雨绮, 唐仕欢, 等. 中药抗 COVID - 19 药效预测分析平台构建及常用中药潜在作用分析 [J]. 中国中药杂志, 2020, 45 (10): 2257-2264.

## 《医学信息学杂志》版权声明

(1) 作者所投稿件无“抄袭”、“剽窃”、“一稿两投或多投”等学术不端行为, 对于署名无异议, 不涉及保密与知识产权的侵权等问题, 文责自负。对于因上述问题引起的一切法律纠纷, 完全由全体署名作者负责, 无需编辑部承担连带责任。(2) 来稿刊用后, 该稿包括印刷出版和电子出版在内的版权、复制权、发行权、汇编权、翻译权及信息网络传播权已经转让给《医学信息学杂志》编辑部。除以纸载体形式出版外, 本刊有权以光盘、网络期刊等其他方式刊登文稿, 本刊已加入万方数据“数字化期刊群”、重庆维普“中文科技期刊数据库”、清华同方“中国期刊全文数据库”、中邮阅读网。(3) 作者著作权使用费与本刊稿酬一次性给付, 不再另行发放。作者如不同意文章入编, 投稿时敬请说明。

《医学信息学杂志》编辑部