

基于大数据应用的医学图书馆服务定位

董瑞玉 冯占英 张晓梅 张 玉

(解放军医学图书馆 北京 100039)

〔摘要〕 阐述大数据在医学领域的应用与瓶颈, 分析大数据应用对医学图书馆挑战, 探讨大数据环境下医学图书馆的服务定位, 包括基于数据整合的资源共享服务、基于读者特定需求的个性化服务、基于数据分析的知识服务等。

〔关键词〕 大数据; 医学图书馆; 服务

〔中图分类号〕 R-056 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2017.01.017

Medical Library Service Location Based on Big Data Application DONG Rui-yu, FENG Zhan-ying, ZHANG Xiao-mei, ZHANG Yu, Medical Library of PLA, Beijing 100039, China

〔Abstract〕 The paper describes the application and bottleneck of big data in the medical field, analyzes the challenges of big data application to medical library, discusses the service location of medical library in the environment of big data, including the resource sharing service based on data integration, personalized service based on special requirements of readers, and knowledge service based on data analysis.

〔Keywords〕 Big data; Medical library; Service

1 引言

在当今科技的快速发展下, 互联网、云服务、数据存储方式等技术日臻成熟。在大数据概念的推广中, 数据信息作为更加实用的资源, 已经在生产发展中呈现更多的实际价值。医学领域同样处于数据爆炸的趋势中, 医学图书馆作为医学行业的信息服务保障机构, 如何运用大数据技术, 调整服务定位, 为读者提供更好的服务, 是图书馆健康长久发展的根本之计。

2 大数据在医学领域的应用与瓶颈

2.1 应用情况

鉴于大数据潜在的巨大影响, 很多国家或国防组织部将大数据视作战略资源, 将大数据提升为国家战略。美国政府将大数据看成是“未来的新石油”。2012 年 3 月美国奥巴马政府宣布“大数据研发计划”, 设立 2 亿美元的启动资金, 希望突破性挖掘生物医药和环境科学等多科学领域大数据资源; 联合国推出的“全球脉动”项目, 希望利用“大数据”预测某些地区的失业率或疾病爆发等现象, 以提前指导援助项目^[1]。国内外医学领域对大数据的应用正产生着日新月异的变化, 大数据被广泛应用在医学领域的各个方面。如在预防医学方

〔修回日期〕 2016-08-26

〔作者简介〕 董瑞玉, 博士, 馆员, 发表论文多篇。

面,利用大数据对各种流行病和传染病进行预测,能够有效预警重大传染性疾病,及时将疫情扼杀于萌芽中。在临床医学方面,通过整合和分析患者试验样品、组织切片、基因芯片等多来源大数据,能够实现个性化医疗和精准医疗,有效提高疾病早期诊断和疑难杂病确诊率,颠覆传统侧重患者个体发病特定原因的治疗技术。通过对致病因素的关联分析,能够对病情发展进行预测,提高诊断准确率,大幅减少误诊误治。在生物医学方面,大数据被广泛应用在基因和蛋白质组学研究、药学研究和个性化医疗方面,使原来由假设驱动生物医学研究正逐步转向数据驱动^[2]。

2.2 瓶颈

2.2.1 给医学领域带来了一定的技术和经济影响

随着信息技术的发展,医学大数据首先在量上不断膨胀,其增长速度已经远远超过计算机存储和计算能力的发展,数据存储技术存在瓶颈问题。二是数据结构越来越复杂,医疗方面主要有临床试验数据、组织切片数据、影像数据、患者资料数据等,生物科研方面主要有试验数据、切片数据、基因芯片数据等,医学大数据用户覆盖范围广,各类数据来源采取的数据采集标准不一,影响数据的流动性和共享。三是部门之间和医院之间信息孤岛现象普遍存在,科研机构间的数据共享名存实亡。

2.2.2 医学大数据的应用环境亟需改善 首先,数据采集是数据有效应用的前提,源数据和数据库的异质性,可能是导致数据矛盾、数据重复利用的重要原因,因此医学数据采集需要有统一的元数据标准。其次,医学大数据的来源纷繁复杂,可能来自不同地区、不同研究团体或者不同医疗机构,这就要求不同医疗单位、科研团体的数据资源整合管理,实现资源共享^[3]。第三,医学数据处理要求较高的结果准确性和较快的处理速度,如个性化医疗,要求具有较高的时效性,而单细胞测序及诊断等,对突变位点和功能模块的鉴别准确性要求较高。第四,有关大数据共享应用方面的政策制度有待于进一步完善。

3 大数据环境对医学图书馆的挑战

3.1 多源数据需要整合集成管理

医学数据种类繁多,主要包括各种不同的基础、临床、预防和生物医学的试验数据和诊疗观察数据,且这些数据资源大多分散于各不同机构,没有统一的平台实施管理,造成同类型科学数据分裂、不同项目之间的科研交流不畅、数据难以共享等问题,从而使得许多研究项目重复进行,导致人力、物力、财力资源的浪费^[4]。经过多年建设发展,我国医学领域已经建立了几百个医学科学资源共享数据库,然而随着医学的不断发展,医学学科越来越细化,新的学科不断涌现,医学类发表文献呈每年递增趋势,尚有一部分数据资源有待于充分开发、整合,建库潜力巨大,在医学科学数据集成、共享方面还有一定的问题亟需解决。科学数据共享是提高科学研究效率的重要基础,构建医学科学数据共享平台是医学图书馆面临的一项重要使命,通过对数据资源进行科学管理,为医学工作者提供信息共享服务的支撑平台,最大限度发挥医学科学数据应有的作用,促进国内外科学数据资源的有效利用,避免低水平研究和重复性研究,节约人力、物力和财力^[5]。

3.2 医学研究要求图书馆提供知识驱动的研究环境

在大数据环境下,随着网络技术和信息技术的不断发展,人们获取知识的方式和途径也在逐步发生改变。医学图书馆的读者群体主要是医学科学研究人员,具有一定的学历及知识水平及较高的信息检索和搜集能力,能够从多渠道获得信息资源,对传统的文献资源和电子资源的依赖程度逐渐降低。读者需要更多的是医学信息数据整合、分析,从而发现新的知识^[6]。读者需求决定图书馆的服务,大数据环境下,医学图书馆以文献服务为主的传统服务模式将面临冲击,为了适应读者需求,图书馆必须调整服务模式,由传统文献服务模式向数字化主导的支持知识驱动的研究服务模式转变。

3.3 馆藏数据可用性有待提高

大数据的可用性概括来说就是每个数据的表达信息不能含有语义错误或者属性相互矛盾,数据能够支持各种不同的查询和分析计算,具体表现为数据的一致性、精确性、完整性、时效性和实体同一性,如某医疗数据库中的数据一致且精确,但遗失某些患者的既往病史,从而存在不完整性,可能导致不正确的诊断甚至严重医疗事故。信息技术的飞速发展,医学领域进入了海量数据时代,随着医学数据资源总量以级数倍增,数据中心的数据管理、存储、计算与处理能力相对于数据增长速度有一定的滞后性,不能完全满足读者大数据服务需求。数据急剧膨胀造成数据资源的价值密度和可用性降低,对有效进行数据挖掘和价值发现产生影响。图书馆在海量数据中的价值发现过程,可能会有较高的成本投入和消耗大量系统资源,多源数据不同的组织结构和采集方式可能产生更高的数据噪声和冗余,导致大数据的价值密度和相关性大幅下降,大数据资源服务的价值转换率、数据可靠性、数据对服务创新的价值、数据对客户特征和服务过程覆盖的全面性等问题,这些都对大数据的实用性产生了一定的影响^[7]。

4 大数据环境下医学图书馆服务定位

4.1 基于读者特定需求的个性化服务

4.1.1 概述 医学图书馆是医学领域信息汇聚中心,在大数据环境下,医学诊疗和研究的改革导致用户需求向多元化和个性化发展。读者阅读需求内容与模式的不确定性增强,读者角色定位由原来的被动接受服务转为决定图书馆服务内容、模式与方法有效性的决定性因素^[8],要求图书馆行业服务由借阅服务、馆藏资源服务等基础被动服务转向依据大数据挖掘探究读者需求,从而有目的、有针对性地满足读者特定需求的个性化服务。

4.1.2 创建客户管理系统 客户管理系统采用类比、聚类分析和用户细分等方法,通过对视频监控器、传感器、管理服务器日志、移动阅读终端等采集的数据进行分析、决策,对读者行为数据进行深

度挖掘,推算出用户潜在需求的数据信息,按照用户需求服务和阅读特点,将读者划分为若干个具有不同服务保障需求组类,为不同组类制定个性化的服务组合。

4.1.3 图书馆和读者双向互动 一方面,读者可以根据图书馆主页推出的服务项目进行个性化定制。另一方面,图书馆根据读者的定制需求,实时采集读者需求动态,以增强读者个性化服务的细粒度为目标。精细化即分析读者个性化服务内容与模式,自动反馈、调整,确保服务系统可依据用户需求变化快速反应,为每一位读者量体裁衣^[10]。为保证采集到低成本、高可用性的数据,图书馆需要对个体读者进行全方位、不间断的全程数据跟踪。通过构建长期行为跟踪模型,来实现对读者注册、借阅预订、阅读行为、满意度反馈、阅读社会关系和信息交流等活动的全程跟踪,达到真正全方位掌握读者阅读需求及动态变化趋势,最终提高图书馆用户个性化服务的效率和读者阅读满意度。

4.1.4 保护读者隐私不受侵犯 图书馆对读者进行个性化服务,需要采集用户行为数据和用户社会关系数据等,在采集这些数据时应该让读者有知情权,在不影响使用的前提下,保护读者数据采集处理过程中的隐私安全性,图书馆应加强有关工作人员职业道德教育与道德规范,避免读者隐私数据的过度挖掘、分析和使用。

4.2 基于数据整合的信息共享服务

4.2.1 概述 随着精准医疗等技术的发展,医学科学数据共享越来越受到重视,早在 20 世纪 80 - 90 年代,美国、欧洲、日本等国家已经率先建立世界级生物数据中心,如美国国家生物技术信息中心(NCBI)、日本 DNA 数据库(DDBJ)、欧洲生物信息研究所(EBI)等,垄断掌握全世界的生物数据和知识资源。近几年,美国又先后投入大量资金用于生物医学数据共享建设,预计到 2020 年投资总额将达到 6.65 亿美元。

4.2.2 统一元数据标准 我国人口众多,具有丰富的生物医学资源,为提高医学科学数据资源的利用率,推动医学科研进步,加速医学科研成果转化

成实用产品,医学图书馆应该在相关政策的支持下,采用标准化分析方法,制定统一的元数据标准,采用数据仓储、开放获取等技术方式,对不同医疗单位、科研团体的数据资源整合管理,实现资源共享。

4.2.3 遵循数据共享保障机制 共享机制是影响数据开放共享的关键因素,国家层面已经陆续出台了一系列关于数据共享的政策和制度,医学图书馆资源数据共享首先要在有关法律和条令的规范下,规划好适合自己的共享机制,主要包括资源整合机制、信息安全保障体系、信息访问机制、奖励激励及配套管理制度。依据相应共享保障机制,与第3方达成协议,把好数据来源关,在尊重知识产权的前提下,对多源数据汇集整理,将数据依据不同来源、不同性质选择不同的共享方式。

4.2.4 用户共享权限设置 为提高管理效益,共享平台可以将用户分为不同等级,根据用户级别不同给予不同的共享服务。如与图书馆直接业务关系或归属同一系统的可给予无偿完全共享,有些用户只能给予部分数据的无偿服务,一些特定数据只能有偿获取^[10]。医学数据共享是一把双刃剑,一方面缓解了信息孤岛、信息烟囱的问题,另一方面用户使用他人的共享数据进行自己的研究,一定程度上促进科学工作者贡献自己的数据,有利于数据共享系统的建设。

4.3 基于数据分析的知识服务

4.3.1 概述 在大数据环境下,开展知识服务是未来图书馆发展的重要方向,知识服务应该以知识创新为中心、以用户为核心,最终目标是实现知识有效增值。医学数据具有海量、类型复杂、增长速度快的特点,对海量数据的搜集、整理和分析挖掘是医学图书馆的服务重点。

4.3.2 建立数据关联 开展知识服务,首先应用大数据技术建立各种数据关联,将海量、无序的数据进行关联融合,如基础和临床医学数据关联、生物医学信息数据和科研诊疗有效结合、科学数据与阅读文献之间的交互操作,以及过滤整合各信息平台的数据,通过数据关联二次发现数据中蕴藏的价值和隐含的关联性。

4.3.3 把握用户需求 通过建立各种资源及学术

搜索引擎、知识服务引擎、用户需求预测引擎等,在对海量数据进行搜索、比较、聚类分析归纳的基础上,实现数据价值的挖掘并对读者需求进行预测,满足不同读者需求差异而提供准确的知识服务。

4.3.4 科技动态预测 医学图书馆知识服务要支持科技创新、围绕重点领域进行未来演变方向及可能性挖掘与预警,实现医学科技态势监测、科技趋势预见等决策支撑服务。

5 结语

本文从大数据技术在医学领域的应用入手,分析大数据环境对医学图书馆的影响,重点探讨大数据环境下医学图书馆的服务定位策略,提出医学图书馆服务重点应从医疗文献服务转向以数据分析和挖掘为重点的知识服务,最终发展成为医疗科研、决策辅助中心的思路,从而为医学专业读者提供更好的服务,对医学图书馆的长远发展具有一定的借鉴意义。

参考文献

- 1 Lancet T. Moving Toward Precision Medicine [J]. Lancet, 2011, 378 (9804): 1678.
- 2 弓孟春,陆亮. 医学大数据研究进展及应用前景 [J]. 医学信息学杂志, 2016, 37 (2): 9-15.
- 3 宁康,陈挺. 生物医学大数据的现状与展望 [J]. 科学通报, 2015, 60 (5/6): 534-546.
- 4 唐源,吴丹. 国外医学科学数据共享政策调查及对我国的启示 [J]. 图书情报工作, 2015, 59 (18): 6-13.
- 5 马良,蔡金华,尹玲,等. 实施科学数据共享提升医学科技竞争力 [J]. 中华医学科研管理杂志, 2004, 17 (1): 33-35.
- 6 饶敏. 大数据时代综合医院图书馆学科化服务理念与模式创新 [J]. 图书馆学刊, 2015, (4): 50-53.
- 7 李艳明,杨亚东,张昭年,等. 精准医学大数据的分析与共享 [J]. 中国医学前沿杂志·电子版, 2015, 7 (16): 4-10.
- 8 马晓亭. 大数据时代图书馆客户关系管理研究 [J]. 图书馆工作与研究, 2014, (6): 49-52.
- 9 陈臣. 基于大数据融合的图书馆读者个性化服务 QOS 保障研究 [J]. 图书馆工作与研究, 2016, (2): 44-47.
- 10 杨友清,陈雅. 科学大数据共享研究: 基于国际科学数据服务平台 [J]. 新世纪图书馆, 2014, (3): 24-28.