

图书馆大数据知识服务生态体系及其动力机制研究^{*}

刘桂锋 卢章平 化 慧

摘 要 大数据技术为图书馆知识服务带来了发展机遇。基于信息生态链理论,阐述大数据管理系统和大数据服务体系,提出“服务 3.0”的大数据服务理念,并构建图书馆大数据知识服务生态体系模型,在此基础上探讨体系构成要素及其动力机制,以期为图书馆更好地开展知识服务提供理论依据与实践参考。图 4。表 1。参考文献 28。

关键词 大数据 数据管理 数据服务 数据科学 数据馆员

Ecological System of Big Data Knowledge Service in the Library and its Dynamic Mechanism

Liu Guifeng Lu Zhangping Hua Hui

Abstract: The emergence of big data technology provides development opportunities for knowledge service in the library. Based on the information ecochain theory, this paper elaborates the big data management system and service system, puts forward the concept of “service 3.0” of big data service, and builds the model of big data knowledge service system in the library. Finally, the components and dynamic mechanism of the system are well discussed in the paper, which will provide theoretical basis and practical reference for the library to carry out knowledge service. 4 figs. 1 tab. 28 refs.

Keywords: Big Data; Data Management; Data Service; Data Science; Data Librarians

大数据技术为图书馆知识服务提供了前所未有的先决条件与发展机遇,引发了知识服务内容(拓宽知识服务的内涵与外延)和知识服务方式(数据清洗和知识咨询服务)的巨大变革^[1]。大数据知识服务引起了国内外诸多学者的浓厚兴趣。当前,大数据知识服务研究一般集中在三个方面。一是基础理论研究,主要包括大数据知识服务的概念、特征、模式、模型等。如,秦晓珠等^[2]认为大数据知识服务概念模型包括数据、知识、资源、能力、服务、过程和任务七个元素;官思发^[3]在深入分析大数据知识服务关键要素的基础上构建了大数据知识服务的概念模型;Begoli等^[4]提出了包含数据收集与分析、系统组织、数据利用三个阶段的基于大数据的数据发现系统

设计原则。二是平台技术研究,主要包括平台体系结构以及配套的关键技术等。如,李晨晖等^[5]认为大数据平台体系的关键技术有数据管理与处理技术、虚拟化接入技术、全生命周期管理技术、终端交互技术等;孙卓^[6]构建了涵盖知识源与过滤层、知识存储层、知识表示与标引层、知识处理与挖掘层、知识检索与反馈层的图书馆知识服务引擎体系;Cai 等^[7]构建了由数据收集、信息处理、知识检索、决策支持四个部分组成的大数据知识服务平台。三是应用领域研究。如大数据知识服务在 Google、Facebook、Twitter、LinkedIn 等搜索引擎或社交媒体中的商业化应用^[8],其中 Google 是运用大数据技术最为成功的公司之一,通过对客户检索行为的大数据分析而推出目标

^{*} 本文系江苏省图书馆学会学术研究课题“大数据轮动的知识服务生态体系构建研究”(项目编号:14YB34)和江苏省博士后科研资助计划“大数据轮动的图书馆学科服务创新研究”(项目编号:1402106C)的研究成果之一。

广告。除此之外,大数据知识服务还广泛应用于物理科学、医疗健康、人工智能、生物、文化、政治、司法鉴定、商业管理等各个领域。国内外图书馆应用大数据技术主要开展数据挖掘、数据分析工作,以及基于用户数据的数据创意服务、文献关联推荐服务等^[9]。以上研究成果分别从基础理论、平台体系与应用领域的视角对大数据知识服务进行了初步的探索,而本研究则以信息生态链理论为基础构建图书馆大数据知识服务生态体系,并提出“服务 3.0”理念,最后详细分析了体系构成要素及其动力机制,以便为图书馆进行大数据知识服务的理论研究和实践探索提供参考。

1 图书馆大数据知识服务生态体系构建

1.1 体系构建的基本原理

德国学者于 1989 年提出“信息生态学”概念,由此开启了该学科的发展历程^[10]。1999 年,美国学者提出“信息生态系统”概念,认为信息生态系统是在特定环境里由人、实践、技术和价值所构成的系统,占据核心地位的是技术支持下的信息主体人,而不是信息技术。信息主体人包括信息生产者、信息分配者、信息传播者和信息分解者^[11]。信息生态系统主要由信息子、信息素、信息场、信息链、信息网、信息域、信息圈等诸要素组成,其中信息链是通过信息的流动沟通无数信息场的通道,是信息生态系统的“灵魂”^[12]。信息生态链包括信息、信息人和信息环境等基本要素,具有空间结构、时序变动和管理等基本特征^[13]。信息生态链的构成主体是不同种类的信息人,其功能实质是实现不同种类信息人之间的信息流转^[14]。

图书馆大数据知识服务系统是一个由信息、信息人和信息环境等要素构成的特色鲜明的信息生态体系。信息指的是图书馆为读者提供服务的过程中产生的各式各样的数据。信息环境指的是直接或间接影响信息人的生存和发展的各种条件的总和,包括外部信息环境和内部信息环境。信息人在图书馆大数据的产生、存储、计

算、分析和决策等环节中承担着信息生产者、信息分解者、信息传播者和信息消费者等不同角色,这四种角色对信息进行处理,并相互作用,环环相扣,构成信息生态体系。研究图书馆大数据知识服务信息生态体系,有利于从微观视角揭示其构成要素与运行机制,有利于从宏观视角提出图书馆面向读者的大数据知识服务一体化解决方案,为图书馆服务在大数据时代的转型发展提供理论基础和实践参考。

1.2 大数据管理系统构成

借鉴数据生命周期理论,并参考大数据管理架构、系统与平台^[15-17]的相关文献,笔者提出大数据管理系统主要由获取、存储、计算、分析和决策五个子系统组成,其中获取是基础条件,存储、计算和分析是关键环节,决策是最终目标。数据获取是大数据管理的源头,必须把好质量关,合理选取数据源以及数据采集、解析、清洗与自动修复、整合与演化的方法。数据质量需要满足精确性、一致性、完整性、同一性、实效性和真实性等条件。存储系统主要解决数据的海量化、快速增长问题,应对不同格式的数据对存储技术提出的挑战。谷歌文件系统(GFS)和 Hadoop 分布式文件系统(HDFS)为大数据的存储奠定了基础。存储系统的基本功能是满足半结构化、非结构化和复杂结构化的海量异构数据的存储。大数据计算是连接大数据技术与应用的桥梁,属于数据密集型计算,对数据吞吐率、性价比和扩展性要求极高。目前出现了多种典型的大数据计算模式以及与之相匹配的计算系统和工具^[18]。大数据分析一般分为实时数据分析和离线数据分析,其中实时数据分析是主流和核心,主要采用流处理模式、批处理模式及两者的结合。大数据分析方法主要有结构化数据分析、文本分析、网络分析、多媒体分析、社会网络分析和移动分析^[19]。决策系统主要是对大数据分析结果的利用与再利用,将分析结果转换成适合目标用户层次的知识服务,为相关人员进行科学决策提供事实依据,核心工作就是数据解释。

大数据管理系统包括数据中心和服务器等硬件、数据存储和数据库等基础软件,以及大数据分析应用软件。云计算是大数据的基础与支撑技术。Hadoop 是目前最为流行的大数据处理平台,主要包括文件系统、数据库和数据处理等功能模块,其它大数据的处理工具大部分是在 Hadoop 基础上进行功能扩展或延伸。许多公司采用 NoSQL 非关系型数据库进行网络应用开发。大数据管理系统面临一系列技术挑战^[19]:一是对分散的、大量的异构数据源的收集与整合;二是能够满足快速检索、可扩展和隐私保护等功能和性能,保证数据集的存储和管理;三是具备建模、可视化、预测和优化功能并可对不同层次大数据进行实时挖掘分析。

1.3 大数据知识服务体系

在大数据时代,图书馆服务呈现出个性化、差异化、人性化和智能化的特征。根据“web3.0”^[20]和“情报3.0”^[21]的基本理念以及图书馆数据来源和分析方法的不同,本文提出“服务3.0”的概念。“服务3.0”具有目标协同化、数据异构化、方法多元化、技术专业化和服务智能化、结果科学化等特征。图书馆大数据服务体系(如图1和表1所示)主要由以文献信息服务为标志的“服务1.0”、以知识服务为标志的“服务2.0”和以智慧服务为标志的“服务3.0”组成。三者之间并不是孤立的,而是相互依赖、相互联系的,如“服务3.0”以“服务1.0”和“服务2.0”为基础,具有包含关系。

“服务1.0”指的是基于读者在接受服务过程中“被动”产生的书目数据、闸机数据和读者借阅记录数据而开展的外借、阅览等文献信息服务,满足读者对图书馆所提供文献信息的基本需求。书目数据主要是指图书馆订购的纸质图书资料和电子资源等所有馆藏资源的总目,能够反映出图书馆馆藏资源的数量、年度变化趋势、纸质与电子资源所占比例、文献类型分布、学科文献结构等。闸机数据主要是指通过图书馆自动门禁闸机管理系统记录的读者入馆数据,可以反映出

馆的总人数、时间变化趋势、读者专业分布等,特别是能够显示读者入馆的高峰与低谷时间段,以便图书馆做好应急对策。借阅数据主要是指读者借阅馆藏文献的记录,以及与之相关的借阅检索、读者荐购、书评等数据。通过借阅数据能够了解借阅次数、热门图书资料、时间变化趋势、学科类别分布、读者专业分布与年级分布,反映读者的借阅行为与阅读规律,为目标读者识别、个性化推荐、阅读推广奠定基础。“服务1.0”的数据类型主要是结构化数据,因此可以采用 Excel、Origin、SPSS 等传统软件自带的回归分析、时间序列分析、图表分析、因子分析、聚类分析和关联分析等统计分析方法。

“服务2.0”指的是图书馆基于各类馆藏科研数据,采用合适的分析方法与工具,为解决用户具体问题而提供的面向知识内容的增值化知识服务,满足用户在知识一体化战略决策、技术研发、科学研究、学科建设等咨询过程中的全方位、多层次信息需求。馆藏科研数据主要是指图书馆收藏的科研人员对科研过程数据进行加工提炼并用于公开发表出版的以论文、图书、专利、科技报告等多种形式存在的科研成果数据,这些数据是科研人员在技术研发、科学研究过程中“主动”产生的。馆藏科研数据主要以纸质形式和数字形式存在,以一次文献为主,二次文献和三次文献为辅,包括期刊论文、研究报告、专利说明书、会议论文、学位论文、技术标准等。图书馆通过对馆藏科研数据的加工、提炼、融合,并进行深度知识挖掘,能够面向用户提供一体化解决方案的决策咨询报告,如战略咨询、产业创新、技术创新、科技查新、成果评估、学科动态、专题定制等。“服务2.0”的数据类型主要是半结构化数据,需要在简单统计分析的基础上,充分深入到内容单元层面,可以采用基于自然语言处理技术、机器学习(深度学习)的文本分析方法。

“服务3.0”指的是基于社交网络、移动互联网和传感器的多源数据与专家智能相结合的智慧服务,用以满足读者与图书馆之间互动式、体验式、参与式的沟通交往的信息需求,多源数据是由

读者和图书馆员之间“互动”产生的数据。社交网络数据是指读者通过图书馆博客、微博、论坛、QQ 交流群、邮件等方式产生的,是能够反映与体现图书馆管理与服务等内容的动态、互动记录。通过社交网络数据能够了解读者利用图书馆的总体概况、关注重点与热点、用户群体分布、读者偏好行为,为图书馆的潜在读者识别、读者行为跟踪、服务品牌铸造、营销体系构建等提供保障。移动互联网数据是指读者通过手机等移动设备借助移动互联网技术进行的图书馆信息查询、浏览、下载、阅读等行为的记录,主要涉及手机图书馆网站、手机短信提醒、微信、App 等方式,便于读者随时随地接受个性化服务,加强图书馆与读者

之间的沟通。传感器数据是指通过有线或无线传感器网络技术对图书馆不同位置的环境和资源进行智能化感知的数据,如书库、机房、阅览室的温湿度、空气质量等环境监测数据,便于为读者提供人性化的阅读体验以及图书馆的图书保存和节能减排。“服务 3.0”的数据类型主要是非结构化数据,首先应针对不同来源渠道的数据选择专门的方法,如互联网数据的网络分析方法、移动数据的移动分析方法、多媒体数据的多媒体分析方法、社交数据的社会网络分析方法等,然后再在统计分析和文本分析方法的基础上,采用具有知识发现、智能计算和专家推荐等高级功能的语义分析方法。

表 1 图书馆数据来源及服务类型比较

数据来源	数据类别	数据特征	获取方式	分析方法	服务类型
书目数据	结构化数据	结构化、数量小	统计信息	统计分析、可视化分析	服务 1.0
闸机数据	结构化数据	结构化、数量小	统计信息	统计分析、预测性分析	服务 1.0
借阅数据	结构化数据	结构化、数量小	统计信息	统计分析、关联分析	服务 1.0
馆藏科研数据	半结构化数据	语义化、信息丰富	网络获取	文本分析、内容分析	服务 2.0
移动互联网数据	非结构化数据	碎片化、人员固定	移动互联网抓取	移动统计分析、挖掘分析	服务 3.0
社交网络数据	非结构化数据	动态化、社会关系	网络抓取	文本挖掘、语义分析	服务 3.0
传感器数据	非结构化数据	动态化、位置固定	传感器数据抓取	智能语义分析、定位分析	服务 3.0

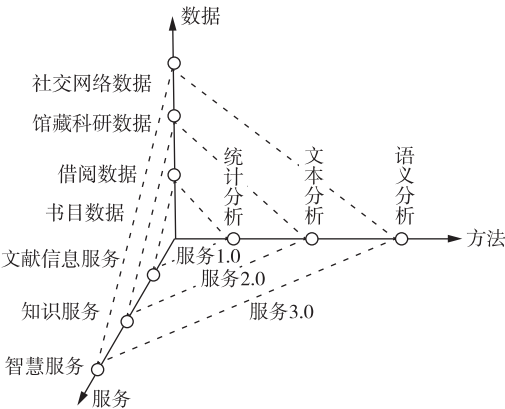


图 1 图书馆大数据知识服务体系

注:不同服务类型的分析方法遵循由初级到高级的变化规律,相互之间并非严格孤立,而是相互依赖、相互联系。

1.4 生态体系模型构建

根据科学性、客观性、有效性和实践性等原则,以及信息生态链的基本原理和图书馆大数据知识服务体系,笔者构建了图书馆大数据知识服务生态体系模型(如图2所示)。生态体系是信息人与内外部环境进行信息流转的一个动态、持续、优化和增值的过程,包括信息人自我完善的内化过程和信息人与外部环境互动的外化过程。该体系是一个耗散结构系统^[22],主要具有开放性、远离平衡态、涨落机制和非线性四个特征。开放性是指科研用户、图书馆员不断与外界进行信息、技术、文化、服务等方面的交流与联系。整个体系在发展过程中也不是孤立的,会受到其它系统及外

部环境的影响。远离平衡态是指图书馆大数据知识服务生态体系的需求方和供给方之间难以达成一致,要么需求方要求过低而供给方轻松满足其需求,要么需求方要求过高而供给方难以满足其需求,造成双方供需之间处于一种非平衡的状态。涨落机制是指在图书馆大数据知识服务生态体系运行过程中,诸如用户需求、信息环境、宏观政策、技术革新等因素会对体系内部信息资源配置的有效性产生影响,使其朝着健康有序的方向发展。非线性是指整个体系不是各个个体的简单加和,而是一个螺旋上升、不断增值的过程。图书馆员对图书馆产生的大数据通过收集、存储、计算和分析,最后为用户提供增值的知识服务,为其决策提供保障。

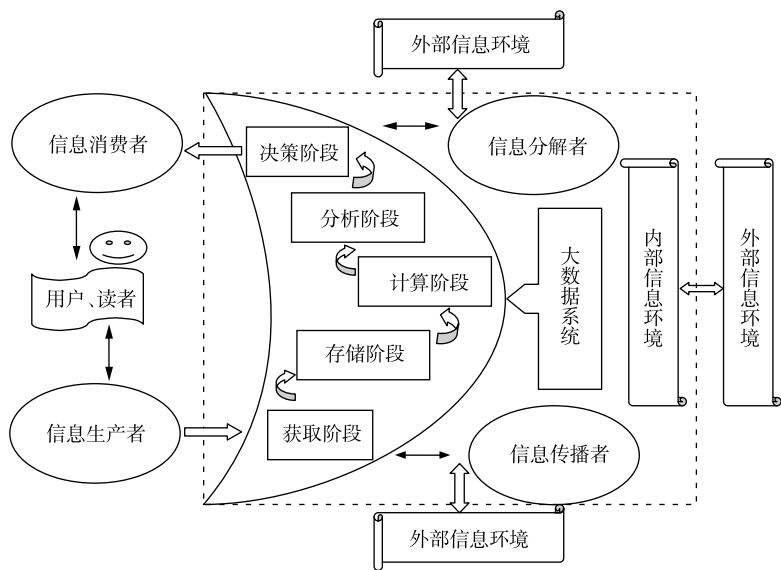


图2 图书馆大数据知识服务生态体系模型

由图2可知,图书馆大数据知识服务生态体系模型主要由信息、信息人和信息环境三部分构成。信息主要是指图书馆产生的各式各样的大数据,如书目数据、闸机数据、借阅数据等。信息环境主要指的是与大数据相关的自然、社会、政治、经济、制度、技术和文化等外部环境和内部环境。信息人主要指的是参与图书馆大数据知识服务过程中的个人或组织。根据信息生态链理论^[23],信息人可以是信息生产者、信息分解者、信

息传播者和信息消费者,信息人的角色和功能是动态变化的,在一定时期或阶段内是可以相互转化的,一个信息人在不同时期可承担不同角色。在该生态体系中,图书馆主要担当信息分解者和信息传播者的角色,并且在数据信息转化为知识产品的过程中发挥着核心作用。数据馆员基于大数据相关的数据科学基本理论,借助图书馆构建的大数据知识服务系统,选择合适的数据分析方法,对数据进行采集、存储、组织、计算、分析等处

理,最后面向用户提供个性化、定制化、智能化、决策化的数据服务。以科研成果数据中的论文数据为例,完成科研成果并撰写论文的科研人员是信息生产者,论文进入图书馆的大数据系统,经过获取、存储、计算、分析等处理阶段,转变为有价值的知识,最后为科研人员提供决策服务,图书馆员担当信息分解者和信息传播者的角色,接收服务的科研人员角色变化为信息消费者,至此形成一个完整的信息生态链,当然还要受到内外部门信息环境的影响。信息生产者的主要功能是生产信息;信息分解者的主要任务是对信息进行知识层面上的加工,以便于利用;信息传播者主要对信息产品进行推介与传递;信息消费者主要是对信息产品的再利用。在图书馆大数据知识服务生态体系中,信息是灵魂,信息环境是支撑,信息人是主体。

2 图书馆大数据知识服务生态体系构成要素及其动力机制

2.1 构成要素

大数据知识服务生态体系主要由人力资源、用户需求、基本理论、数据来源、技术系统、方法应用、服务决策 7 个要素组成(如图 3 所示),各要素之间相互作用、相互联系。

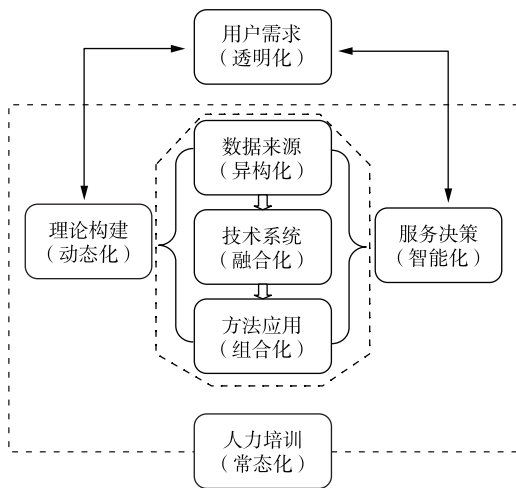


图3 图书馆大数据知识服务生态体系构成要素

人力资源贯穿于生态体系的全过程,在体系设计、运行、监管和评估等环节起着决定性作用,是决定体系运行成败的关键性因素。大数据时代,数据馆员需要具备扎实的基础知识和过硬的基本技能,以保证在体系运行的各个环节做到游刃有余。因此,图书馆应加强顶层设计,统筹规划,制定《数据馆员培训制度》,力争对数据馆员的教育培训做到常态化。

用户需求是服务实现的入口,只有全面了解用户需求的特点,才能提供针对性和个性化的增值服务。大数据知识服务中对用户需求的挖掘应具有透明化特点,图书馆不能再像以前那样仅采取调查问卷、访谈、观察等直接手段了解用户需求,而应在实际调研的基础上,通过大数据手段动态性、实时性地分析用户的信息活动痕迹,理性地发现用户需求及其行为规律。

基本理论是服务实现的源泉,用于指导服务实践的合理实施。图书馆知识服务的基本理论体系主要由原理体系、技术体系、方法体系和应用体系等部分组成。原理体系主要指大数据、数据管理、知识服务、数据生命周期等原理;技术体系主要指图书馆围绕大数据的获取、存储、计算、分析及决策等阶段的关键技术及阶段之间的技术实现;方法体系指的是图书馆采取的由统计分析、文本分析、语义分析等组成的数据分析方法;应用体系指的是图书馆将大数据知识产品和服务应用到不同领域。

数据来源是服务实现的基础,获取与存储全面、多样和动态的数据具有极大的挑战性。图书馆大数据来源具有典型的异构化特点,包括了书目数据、闸机数据和读者借阅记录数据等结构化数据,教学数据、科研过程数据和科研成果数据等半结构化数据,以及社交网络、移动互联网和传感器数据等非结构化数据。异构数据的获取、清洗、解析、融合与存储是至关重要的。

技术系统是服务实现的保障,是图书馆依据数据管理全生命周期面向知识服务的软硬一体优化集成的大数据技术解决方案。图书馆大数据知识服务技术系统具有融合化特点,即技术系统

主要探讨获取、存储、计算、分析及决策等阶段的关键技术及阶段之间的技术实现。图书馆需要借鉴云计算、物联网、移动互联网等信息技术,在学校机构知识库的基础上,深化数据存储功能,扩展数据获取计算、分析及决策功能,应用数据获取技术、数据组织技术、数据存储技术、智能分析技术和智能决策技术,实现数据管理基础设施等硬件和数据库等软件的高效整合与优化。

方法应用是服务实现的关键,是连接服务资源与服务产品的桥梁。图书馆大数据的计算与分析需要满足实时、广度、深度和多维度分析的要求,因此对分析方法提出了新的更高要求,统计分析、文本分析和社会网络分析等传统的分析方法逐渐向网络分析、多媒体分析和移动分析等知识发现、智能计算、语义分析和专家推荐等高级分析方法转变,以及多种方法之间的组合。数据馆员需要对所有分析方法从方法类型、方法原理、方法组合、方法应用等方面进行深度梳理,构建适合不同数据类型的方法体系。

服务决策是图书馆为用户提供的终端服务产品,是对知识服务效果的有效反馈和直接检验。图书馆大数据知识服务的方式、途径、模式等发生了颠覆性的变化,数据馆员需要针对用户信息需求的个性化、人性化和动态化等特点,多层次、全方位、立体化地构建智能化的大数据知识服务模式体系。该模式体系以用户的个性化需求为中心,以图书馆大数据科学分析结果为依托,为用户提供具备知识发现、创新预测、智能提升、科学决策等功能的个性化智能服务。

2.2 动力机制

大数据知识服务生态体系运行机制主要有创新动力机制、需求导向机制、合作协调机制、机构监督机制、信息反馈机制、风险承担机制等,其中创新动力机制是体系构建与运行的原动力,并且起着基础性的决定作用。动力机制主要源于“推动力”和“拉动力”两种力量^[24]。“推动力”是图书馆大数据知识服务发展的基础与保证,“拉动力”是方向与目标。

“推动力”包括直接推动力的“行业竞争加剧”和间接推动力的“技术手段先进”。“行业竞争加剧”主要是指受搜索引擎的影响,用户产生信息需求时的第一反应就是借助搜索引擎来解决问题,导致图书馆的文献保障功能被削弱。因此,图书馆为应对竞争,提出错位发展的理念,积极利用搜索引擎的用户浏览行为特征,为其提供有针对性、高附加值的知识服务。“技术手段先进”是指云计算、物联网、移动互联网等信息技术,Hadoop、MapReduce 等大数据技术,以及数据获取、存储、计算、分析及决策等数据管理技术发展迅速,为知识服务提供了强有力的技术保障。知识服务属于智力密集型行业,需要依赖先进技术手段进行大量数据的处理与分析。

“拉动力”包括内在拉动力的“用户需求提高”和外在拉动力的“先进经验示范”。“用户需求提高”是指随着网络技术的发展以及用户信息素养水平的提高,用户信息需求的层次和水平不断向个性化、专业化、集成化、精准化和高端化过渡,倒逼图书馆知识服务必须转型升级才能跟得上时代要求。“典型经验示范”主要指大数据技术已经被成功运用在电视媒体、社交网络、医疗保险、体育比赛、电子邮件、零售行业等领域,如亚马逊 1/3 的销售额来自于个性化推荐系统^[25]。互联网公司在应用大数据技术方面位居前列,如 IBM 开发了一套复杂的预测模型,完成了电动汽车动力与电力供应系统的预测。这些典型案例的成功做法为图书馆提供了有益的启示与借鉴。

因此,图书馆只有满足用户需求,借助典型经验,运用先进技术,才能在激烈的行业竞争中抢得先机。图书馆只有不断地进行技术创新、服务创新、管理创新和制度创新,大数据知识服务生态体系的运行才会朝着正确的方向不断前进。

3 总结与展望

大数据时代,图书馆知识服务生态体系基本形成。数据是源头,技术是保障,人员是关键,服务是目标。大数据时代图书馆的数据数量巨大、数据种类多元、数据速度加快、数据内涵丰富;大数据

技术以及云计算、物联网、传感网、移动互联网等信息技术的蓬勃发展为图书馆知识服务生态圈的运行源源不断地提供技术保障;人员不仅包括处于核心地位的图书馆员,还包括科研人员、期刊编辑人员、数据库加工人员、信息技术人员、情报分析人员、科研管理人员等;面向不同信息需求的用户提供人性化、个性化和智能化的服务是图书馆服务的最高目标。本文从信息生态链的视角出发,构建了大数据时代图书馆知识服务生态体系,并深入剖析了图书馆“服务 3.0”的服务体系,最后论述了体系的构成要素及其动力机制。虽然图书馆离真正的大数据分析服务还有一段距离,尤其是在数据源融合、技术体系构建、分析方法选择和服务结果决策等方面,但是图书馆进行大数据知识服务的理论基础和实践探索已经开始。

“数据科学是理论根基、数据馆员是核心角色、数据服务是实践价值”的大数据科学观是图书馆未来发展的主要方向,数据科学指导数据服务,数据服务反哺数据科学,数据馆员则是实现数据科学作用于数据服务,数据服务反作用于数据科学的中间桥梁(如图 4 所示)。



图 4 图书馆大数据科学框架

数据科学为数据馆员开展数据服务提供强有力的理论源泉。“数据科学”一词早在上世纪 60 年代就已出现^[26],2002 年《数据科学》杂志创刊,而近来的大数据技术强有力地推动了数据科学的繁荣。大数据时代图书馆数据科学的基本原理、基本方法、应用领域等雏形基本形成。数据科学是由数学、统计学、计算机科学、管理学和图书情报学等多学科组成的综合性学科,主要包括理论体系、技术体系、方法体系和应用体系。理论体系涵盖大数据理论、科研生命周期理论、信息生态理论、服务理论等。技术体系包含网络技术、云计算技术、物联网技术、传感网技术、移动互联网技术

等信息技术,以及涉及数据存储、组织、共享、引用、分析、利用与再利用的数据管理技术。方法体系主要是由统计分析、预测分析、文本挖掘分析、多媒体分析、网络分析、社会网络可视化分析和移动分析等方法组合而成。应用体系是数据科学理论在面向不同服务对象的个性化需求时图书馆“服务 1.0”“服务 2.0”和“服务 3.0”等集成的领域。

数据馆员在将数据科学运用到数据服务的过程中发挥着关键性作用,因此对其角色定位和职能实现提出了更高、更新的要求。数据馆员是在原有学科馆员基础上的延伸和提升,因此只有不断学习和实践才能适应岗位职责要求。国外主要通过数据监管专业教育认证项目、数据监管课程、数据监管继续教育等三种方式来培训数据馆员,以便胜任数据资源建设、数据管理服务与推广等岗位的职责^[27]。数据馆员角色定位于基于数据生命周期的数据管理、数据管理咨询和数据管理培训,同时需要具备扎实的专业基础知识、基本技能与专业技能等数据素养^[28]。专业基础知识是指要掌握与数据科学相关的所有背景资料,基本技能包括策划与组织、领导与管理、合作与交际等能力,专业技能包括网络信息技术、数据管理技术等工具的使用与开发。

数据服务是数据科学在图书馆服务领域的具体实践,是大数据时代图书馆服务新的落脚点和生长点。数据服务的内涵和外延非常丰富,而且一直处于动态更新中,包括服务原理、服务机制、服务内容、服务模型、服务模式、服务体系、服务实践、服务政策、服务评价等。欧美高校图书馆在数据管理服务方面的探索由来已久,并且在服务内容与 service 实践等方面已经取得一定成果,为我国图书馆数据服务的理论建设与实践探索提供了借鉴与参考。

综上所述,大数据时代图书馆知识服务的现实条件基本具备,只有数据科学作为肥沃的土壤,数据馆员作为充足的水分、新鲜的空气和温暖的阳光,两者为数据服务源源不断地提供养料,数据服务这颗种子才能生根、发芽、开花、结果,大数据时代图书馆知识服务生态体系才能真正名副其实。

参考文献

- 1 东方.大数据影响下的知识服务应用模式之变革[J]. 新世纪图书馆, 2015 (6).
 - 2 秦晓珠,等. 大数据知识服务的内涵, 典型特征及概念模型 [J]. 情报资料工作, 2013 (2).
 - 3 官思发. 大数据知识服务关键要素与实现模型研究[J]. 图书馆论坛, 2015(6).
 - 4 Begoli E, Horey J. Design principles for effective knowledge discovery from big data [C]//Software Architecture (WICSA) and European Conference on Software Architecture (ECSA), 2012 joint working IEEE/IFIP conference on IEEE. Washington: IEEE Computer Society, 2012: 215-218.
 - 5 李晨晖,等. 大数据知识服务平台构建关键技术研究[J]. 情报资料工作, 2013(2).
 - 6 孙卓. 基于大数据构建图书馆知识服务引擎研究 [J]. 图书馆学研究, 2013(18).
 - 7 Cai D, et al. The knowledge service project in the era of big data [C]//Big Data (BigData Congress), 2013 IEEE International Congress on IEEE. Washington: IEEE Computer Society, 2013: 429-430.
 - 8 Kosala R, Kumaradajja R. Big Data Analytics for Business [M]//Khosrow-Pour M (Ed.). Encyclopedia of Information Science and Technology. Hershey: IGI Global, 2015:355-362.
 - 9 沈杰. 大数据与图书馆信息服务工作的变革 [J]. 图书馆, 2015(9).
 - 10 Capurro R. Towards an information ecology [C]//Wormell. Information Quality, Definitions and Dimensions. London: Taylor Graham, 1990: 122-139.
 - 11 Nardi B A, O'Day V. Information ecologies: Using technology with heart [M]. Cambridge: MIT Press, 1999:1-262.
 - 12 李美娣. 信息生态系统的剖析[J]. 情报杂志, 1998(4).
 - 13 韩刚, 覃正. 信息生态链: 一个理论框架 [J]. 情报理论与实践, 2007 (1).
 - 14 娄策群. 信息生态链: 概念, 本质和类型 [J]. 图书情报工作, 2007 (9).
 - 15 郭自宽,等. 大数据生态系统在图书馆中的应用 [J]. 情报资料工作, 2013(2).
 - 16 工业和信息化部电信研究院. 大数据白皮书[R]. 北京: 工业和信息化部电信研究院, 2014:1-26.
 - 17 孟小峰, 慈祥. 大数据管理: 概念, 技术与挑战[J]. 计算机研究与发展, 2013 (1).
 - 18 中国计算机学会大数据专家委员会. 中国大数据技术与产业发展白皮书[R]. 北京: 中国计算机学会大数据专家委员会, 2013:1-114.
 - 19 Hu H, et al. Towards Scalable Systems for Big Data Analytics: A Technology Tutorial [J]. IEEE Access, 2014(2):652-687.
 - 20 刘桂锋. Web3.0 及其在图书馆应用的文献综合分析[J]. 图书情报研究, 2011(4).
 - 21 吴晨生,等. 情报服务迈向 3.0 时代[J]. 情报理论与实践, 2015(9).
 - 22 吴婵君. 基于耗散结构的中小企业金融服务体系运行机制研究[J]. 浙江树人大学学报, 2014 (6).
 - 23 娄策群. 信息生态位理论探讨[J]. 图书情报知识, 2006(5).
 - 24 毕强, 白云峰. 金融领域信用信息服务作用机理与运行机制研究[J]. 情报资料工作, 2011 (3).
 - 25 中国大数据 [DB/OL]. [2015-10-23]. <http://www.thebigdata.cn/>.
 - 26 汪小帆. 数据科学与社会网络: 大数据, 小世界[J]. 科学与社会, 2014 (1).
 - 27 高珊, 卢志国. 国外数据馆员的能力需求与职业教育研究[J]. 图书馆, 2015 (2).
 - 28 魏来, 高希然. 大数据背景下高校数据馆员的角色定位[J]. 情报资料工作, 2015 (5).
- (刘桂锋 副研究馆员 江苏大学科技信息研究所, 卢章平 教授 江苏大学图书馆, 化慧 江苏大学科技信息研究所图书情报与档案管理专业 2013 级硕士研究生)

收稿日期:2015-10-27