

大数据在高校图书馆信息资源建设中的应用探析

徐春艳, 许丽丽, 洪艳霞

(黑龙江大学图书馆, 黑龙江 哈尔滨 150080)

摘 要: 结合大数据相关理论及高校图书馆信息资源建设面临的问题和挑战, 探讨大数据应用于高校图书馆信息资源建设的价值与方法。高校图书馆利用大数据进行信息资源建设, 应注重选择适当的平台与工具、建立健全相关规章制度、促进馆员数据技能发展。

关键词: 大数据; 高校图书馆; 信息资源建设

中图分类号: G253

文献标识码: A

文章编号: 1002-1248 (2017) 02-0027-03

Application of big data in University Library's Information Resource Construction

XU Chunyan, XU Lili, HONG Yanxia

(Library of Heilongjiang University, Harbin 150080, China)

Abstract: Combining with the problems and challenges of university library's information resource construction, as well as big data theory, this paper discussed the value and methods of applying big data in university library's information resource construction. During this process, university library should pay attention to select appropriate platform and tools, establish the relevant rules and regulations, and promote the development of librarians' data skills.

Keywords: big data; university library; information resource construction

1 大数据概述

大数据是一个比较抽象的概念, 目前国内外并无统一的定义。人们通常将其定义为所涉及的数据量巨大, 无法通过常规软件工具, 在可承受的时间范围内对其内容进行捕捉、存储、处理和分析的数据集合。大数据的基本特征可以归纳为 4 个“V”, 即数据规模大 (Volume)、数据种类多 (Variety)、生成速度快 (Velocity)、价值密度低 (Value)。

大数据既是一种丰富数据的概括, 也是一种先进的技术, 更是一种发现事物发展本质的思维方式, 可以从不同的角度加以理解。从大数据本身的角度看, 大数据是海量、多源、异构、动态的大型数据集, 其价值密度相对较低, 需要快速、持续的分析处理和挖掘才能得到真正有用的信息, 形成用户价值。从技术的角度看, 存储、计算和分析技术与工具的发展, 尤其是云计算的出现为海量数据的存储和运算提供了有

效的解决方案, 使对大数据的深入挖掘成为可能。从思维变革的角度看, 人们在大数据时代对待数据的思维方式将发生 3 个转变: 一是人们处理的数据从样本数据变成全部数据; 二是人们不得不接受数据的混杂性, 而放弃对精确性的追求; 三是人们通过对大数据的处理, 放弃对因果关系的渴求, 转而关注相关关系^[1]。

对于作为社会信息中心和知识存储中心的图书馆而言, 大数据技术和思维则要求其拓展资源建设与服务的理念, 更加重视用户数据及信息利用行为, 从数据的视角去发现用户的潜在需求, 进而创新用户服务的内容和方式。

2 高校图书馆信息资源建设面临的新挑战

信息技术的飞速发展, 引发了信息资源获取渠道、图书馆用户需求和上游产业链的一系列变化, 给高校图书馆的信息资源建设带来多重挑战。

收稿日期: 2016-07-18

基金项目: 黑龙江省高校图工委项目“大数据环境下高校图书馆信息资源建设策略研究”(项目编号: 2015-A-093); 2014 年黑龙江省哲学社会科学规划项目“黑龙江省高校图书馆数字资源区域性协作建设与管理研究”(项目编号: 14D042)

作者简介: 徐春艳 (1966-), 女, 黑龙江大学图书馆, 副研究馆员, 已发表论文多篇; 许丽丽 (1975-), 女, 黑龙江大学图书馆, 馆员; 洪艳霞 (1962-), 女, 黑龙江大学图书馆, 副研究馆员。

2.1 信息资源获取渠道的多元化对高校图书馆的传统角色形成挑战

信息技术的发展推动了高等教育和学术研究的转型,开放获取期刊、机构知识库等开放获取资源日益增多,开放教育资源和原生数字资源不断涌现。人们的信息资源获取渠道呈现多元化的趋势,可以利用各种搜索工具,通过互联网、移动互联网获取丰富的学术信息,图书馆不再是其资源获取的唯一途径。高校图书馆作为机构资源中心的优势和地位在信息资源多渠道获取的背景下日益弱化,各种相关调查亦显示高校图书馆的用户访问量和纸质图书借阅量呈现逐年下滑的趋势^[2],高校图书馆的资源建设前景不容乐观。因此,高校图书馆需要扩展其传统的资源保存和管理角色,建立更全面的馆藏发展战略,广泛收集新兴的开放获取资源、开放学习资源及其它免费的网络教学和学术资源并实现各种资源的有效整合,从而提升高校图书馆资源的受关注度、被利用率和整体竞争力。

2.2 用户信息资源需求的多样化对高校图书馆的资源配置形成挑战

信息技术的发展引发了图书馆用户信息利用行为的变化,使其对数字资源的需求日益增长。相应地,国内外高校图书馆采购经费越来越偏重于数字资源,电子期刊和电子图书的经费比例呈现逐年增加的态势。而不同学科的教师和学生在使用纸质资源和数字资源方面还存在着显著差异,人文学科侧重于使用纸本,而工科和医学学科侧重于使用数字资源。但各学科领域的基础文献,如基础理论方面的学术专著仍倾向于使用纸本图书^[2]。与此同时,国内外高校图书馆普遍面临文献采购经费减少和文献价格不断上涨的困境,高校图书馆需要考虑不同学科用户资源需求的差异性,以有限的经费满足用户的多样化需求和学科发展的实际需要,如何合理配置资源比例、优化馆藏结构,实现不同学科、不同类型资源的协调发展,已成为高校图书馆面临的重要挑战。

2.3 新型采购模式的不断涌现对高校图书馆的业务流程形成挑战

出版、发行等图书馆上游产业链也出现了许多前所未有的变化。文献资源出版商、供应商改变原有的资源销售模式,提供按量付费、按需印制等新的服务方式,需求驱动采购(Demand Driven Acquisition, DDA)、读者决策采购(Patron Driven Acquisition, PDA)等新型采购模式不断发展。例如,全球知名学术出版商Wiley近年推出循证采购(evidence-based collections)电子书销售模式,图书馆需要预先支付出版商全部电子

书(或其分库)一年的访问费用,然后可以根据读者的实际利用情况,从中选择图书,将其添加到馆藏,获得该电子书的所有权,但必须达到事先约定的最低数量^[3]。这些新型采购模式,对高校图书馆资源建设的既有流程形成挑战,高校图书馆需要调整业务流程,建立与之相适应的新的资源采集标准和馆藏评估方法。

3 大数据在高校图书馆信息资源建设中的应用价值

大数据技术为高校图书馆深入分析和挖掘用户需求提供了可能,有助于高校图书馆基于数据分析制定采购策略,建立以用户需求为导向的资源建设模式,科学规划资源建设与发展。大数据在高校图书馆资源建设中的具体应用价值可概括为以下三方面。

3.1 改进采购流程

高校图书馆的资源建设工作,主要由采访馆员依据文献采选原则及年度采选计划,结合文献价格因素,从大量书目或书源信息中筛选出适合馆藏的文献,并订购入藏。对于用户需求的了解和判断,通常是咨询相关学科专家意见或由其直接划定目录,以及通过读者荐购方式获取用户需求信息。所采购图书的品质更多地依赖于相关工作经验。而大数据技术与工具的运用,可以使高校图书馆资源选择、采购和处理的过程转变为由数据分析所支持的进程,在整个馆藏管理周期全面考虑用户需求,进而使图书馆资源更加符合用户兴趣,将图书馆采购由个人经验模式转向经由数据分析进行决策参考的模式。

3.2 优化资源配置

网络的发展和普及催生了越来越多的用户信息行为数据。除了图书流通、电子资源检索、浏览、下载等用户资源利用数据以外,用户与图书馆之间的互动数据,如用户对图书馆资源的评价、评分、推荐等活动的数据,都能够在一定程度上反映用户的兴趣、需求和使用意愿。图书馆博客、微博、微信等社会化媒体的应用发展,也积累了大量的用户信息行为和需求数据。高校图书馆利用大数据技术,广泛汇集上述不同来源的用户数据,通过对不同学科、不同层次用户的信息行为、阅读兴趣、借阅需求等各种数据的深入挖掘和分析,可以及时掌握各学科用户的资源需求特征,并根据其实际需求进行资源建设,从而实现采购经费的合理分配,优化资源配置和馆藏结构,增强资源供给的针对性。

3.3 完善资源评价体系

资源的选择与评估是图书馆资源建设工作必不可少的重要环节,科学有效的资源评价体系是合理配置

资源的基础。图书馆传统的资源评价通常采用定性与定量相结合的方法,主要衡量指标有图书借阅率、电子资源的访问、检索、全文下载量、资源被引频次以及经济效益指标等。新的学术交流模式及社会化媒体的发展对图书馆的资源评价产生了很大影响,一方面促使评价对象扩展到机构知识库的内容及其它开放获取学术资源;另一方面催生了社会化的资源评价方式,如替代计量(Altmetrics),就是一种基于社会网络的学术影响力评价方法^[4]。相应地,高校图书馆也需要变革资源评价的内容与方法,以指导新形势下的资源建设。高校图书馆利用大数据工具,能够对来自图书馆系统和社交网络等不同渠道的异构数据加以整合,综合运用各种传统的资源评价指标和新兴评估方法,实现更全面的数据分析,促进资源评价效果的提升。

4 大数据在高校图书馆信息资源建设中的应用建议

4.1 选择适当的平台与工具

高校图书馆要实现海量且复杂的非结构化数据的获取、存储和处理,需要以高性能的硬件设备和软件平台为支撑。中国高校图书馆自20世纪90年代以来,不断扩充服务器,开发或引进各类信息系统,如图书馆自动化管理系统、资源的管理、发现与获取系统^[5],已具备一定的数据存储能力,但在数据统计分析方面,主要还是采用传统的数据分析方法和工具,尚不能满足对非结构化数据深入挖掘的需求。

Hadoop是目前应用最广泛的分布式系统架构,主要用基于计算机集群环境的大规模数据分布式处理,具有可靠性高、扩展性好、运行速度快等优势。Hadoop在图书馆界已有不少应用,例如,新加坡国家图书馆管理局(Singapore National Library Board, NLB)就建立了Hadoop分布式系统集群,作为大数据分析的基础平台,解决了数据存储与计算设备扩展性的问题。在此基础上,NLB利用文本分析软件和协同过滤技术,通过数据分析预测用户需求,并根据需求情况进行馆藏规划^[6]。中国高校图书馆可借鉴国外经验,充分利用已有硬件设施,争取技术企业支持,搭建大数据分析平台并选择适宜的大数据分析和挖掘工具,为图书馆应用大数据进行资源建设与服务提供基础。

4.2 建立健全相关规章制度

高校图书馆应用大数据进行信息资源建设,是一种全新的尝试,需要建立、健全相关规章制度来进行规范和指导。高校图书馆应结合图书馆资源选择、评估等各个环节,制定大数据收集、存储、挖掘和

分析的业务规范,为相关工作人员的业务操作提供指引。高校图书馆还应对资源评价标准进行修改,增设馆藏资源大数据分析的相关指标。此外,由于广泛收集各种来源的用户数据,且对用户身份、借阅需求、阅读兴趣、信息行为等多种数据进行关联挖掘和分析,会更多地涉及用户个人信息和隐私,因此高校图书馆也需要建立相应保密规范及隐私保护制度,加强对用户隐私的保护。

4.3 促进馆员数据技能发展

高校图书馆应用大数据进行资源建设和服务,对馆员的数据技能提出了更高要求。馆员需要具备一定的数据分析能力,掌握大数据平台及相关工具的使用。目前,中国高校图书馆馆员的数据技能普遍缺乏,因此,需要积极引进数据专业人才,同时大力促进现有员工的数据技能发展。具体来说,可以邀请专家为员工举办大数据知识讲座,搜集相关学习资料并组织员工学习和研讨,选派员工参加针对图书馆的大数据技术培训班,以及为员工提供其它数据技能培训和交流的机会。高校图书馆还应促进数据专业人员与采访馆员、学科馆员等工作人员合作,实现馆员之间知识和技能的互补,也使馆员在工作实践中提升自身的数据技能。

综上所述,大数据思维与技术的应用有助于高校图书馆优化采购流程和资源配置,建立以用户需求为导向的资源建设模式。高校图书馆利用大数据进行信息资源建设,需要从技术、管理、人才建设等多个方面共同推进,才能取得良好的效果。

参考文献:

- [1] (美)维克托·迈尔·舍恩伯格.大数据时代:生活、工作与思维的大变革[M].杭州:浙江人民出版社,2013.
- [2] 王晶晶,朱本军,肖珑.演变中的学术环境和图书馆资源建设新方向——第三届中美高校图书馆合作发展论坛会议综述[J].大学图书馆学报,2015,(5):5-11.
- [3] Association of College and Research Libraries. Environmental Scan 2015[EB/OL]. [2016-06-06]. <http://www.ala.org/acrl/sites/ala.org/acrl/files/content/publications/whitepapers/EnvironmentalScan15.pdf>.
- [4] Wikipedia. Altmetrics[EB/OL]. [2016-06-06].<https://en.wikipedia.org/wiki/Altmetrics>.
- [5] 包冬梅,范颖捷,李鸣.高校图书馆数据治理及其框架[J].图书情报工作,2015,(18):134-141.
- [6] Tao Ai Lei. Singapore Library Mines Bigdata [EB/OL].[2016-06-06]. http://bigdataanalytics.my/downloads/BI_Asean_Sept_2014.pdf.