

大数据时代背景下高校图书馆的服务变革探析

郑红月, 蒋丽平, 李如辉

浙江中医药大学图书馆, 浙江 杭州 310053

摘要: 鉴于图书馆界对“大数据”基本未作出应有反应现状, 本研究就“大数据”概念与历史发展及其由来、“大数据”战略价值、“大数据”时代高校图书馆面临的机遇与挑战以及对策等作一探讨, 以冀引起业界对“大数据”的关注。

关键词: 大数据; 高校图书馆; 服务变革

doi: 10.3969/j.issn.2095-5707.2014.03.003

Analysis of University Library Service Reform in the Era of Big Data

Zheng Hongyue, Jiang Liping, Li Ruhui

(Library of Zhejiang Chinese Medical University, Hangzhou Zhejiang 310053, China)

Abstract: Since library community basically has not yet made proper response to Big Data, this article elaborated the concept, the historical development, the origin, the strategic value of Big Data, and the opportunities and challenges university libraries encountered in the era of Big Data, in order to draw public attention.

Key words: Big Data; university library; service reform

近十年来, 信息技术热点的形成日益呈现出“加速”趋势。目前, 继数字图书馆、移动图书馆、物联网、云计算等热点之后, “大数据”(Big Data)无疑已登居IT热词之榜首。尽管诸多学者已经或正在努力地描绘着“大数据”时代的崭新图景——“大数据”对世界的广泛而深刻的变革, 尽管以信息服务为核心职能的图书馆在这场变革中难以置身其外, 但毋庸讳言, 迄今为止还没有引起图书馆界足够的重视——图书馆界基本上还没有作出反应。有鉴于此, 本文试就“大数据”概念与历史发展及其由来、“大数据”战略价值、“大数据”时代高校图书馆面临的机遇与挑战以及对策等作一探讨, 以冀引起业界对“大数据”的关注。

1 “大数据”概况

1.1 “大数据”概念及其历史发展

一般认为, “大数据”作为一个相对成熟、相对定型、在核心内涵认识上达成共识的概念, 见于美国信息存储资讯科技公司(EMC)在2011年5月召开的“EMC World 2011”大会上, 美国麦肯锡咨询公司(Mckinsey & Company)在其所发布的研究报告《大数据: 创新、竞争和生产力的下一个前沿》中用“大数据”这一概念表述“创新、竞争和生产力”的下一个前沿^[1]。倘若着眼于一般用词, 其历史则可以回

溯到上个世纪末, 1998年, 美国《科学》(Science)杂志在介绍某计算机软件时就曾提及“大数据”一词。与麦肯锡咨询公司报告以及近年来所讨论的“大数据”概念最为接近的文献则出现在2008年英国《自然》(Nature)杂志推出的“大数据专刊”。

国内外知名媒体对“大数据”的专题报道, 尽管对概念的理解不尽相同, 但在定义上又有着基本共识——数据数量的大小、通信及其便利并非“大数据”的本质, “大数据”的特质在于在海量数据中快速获取所需信息、学到知识^[2]。维克托·迈尔·舍恩伯格(Viktor Mayer-Schönberger)教授在其著作《大数据时代: 生活、工作与思维的大变革》中认为, “大数据”并非一个明确的概念, “大数据是人们在大规模数据的基础上可以做到的事情, 而这些事情在小规模数据的基础上是无法完成的; 大数据是人们获得新的认知、创造新的价值的源泉; 大数据还是改变市场、组织机构, 以及政府与公民关系的方法。”^[3]

1.2 “大数据”的由来

“大数据”衍生于信息数据数量上的持续“加速”增长。美国著名市场研究公司IDC研究显示, 全球信息数据量以年均50%的速率迅猛增长, 2011年全球新增数据量高达1.8 ZB(1 ZB=10亿TB)^[4]。美国南加州大学的学者在《自然》杂志中撰文指出, 人类信息储存量每18个月翻一番^[5]。数字资源构成的变化乃信息数据数量剧增的最为根本的原因。随着智能手机、穿戴式终端等移动终端的普及, 以及社交网络、

基金项目: 浙江中医药大学校级课题(2012ZZ09)

第一作者: 郑红月, 副研究馆员, 研究方向: 期刊评价与研究。E-mail: 1074194096@qq.com

移动互联网、物联网、信息化应用、传感器、摄像头等方面的快速发展,原生的(Born-digital resources)“非结构化数据”剧增,并成为“数据资源的主体”,目前,已占数据总量的80%以上^[4]。海量的、杂乱无章的数据中蕴藏的巨大价值引发了对于数据处理、分析、挖掘的巨大需求,这样便催生了“大数据”概念与相关技术。

2 “大数据”的价值

2.1 “大数据”的战略意义

“大数据的真实价值就像漂浮在海洋中的冰山,第一眼只能看到冰山的一角,绝大部分都隐藏在表面之下。”^[3]“大数据”的价值有如浪里淘沙,却又弥足珍贵,目前已经得到了西方发达国家国家战略高度的重视,2012年3月29日,美国政府宣布推出“大数据研究和发展计划”(Big Data Research Development Initiative),斥资两亿美元研发采集、组织和分析大数据的工具及技术^[6]。

就高校图书馆而言,如何利用大数据技术服务于高校图书馆发展策略的制定、业务的转型拓展与提升等,已经成为高校图书馆行业层面的战略课题。

2.2 思维范式变革

“大数据”最具颠覆性的影响在于其对思维范式的变革,即抛弃对样本、精确性、因果关系的执迷,转而接受整体、混沌、相关关系,更好地知觉当下以及正在发生的未来。对高校图书馆而言,提高对数据资源的重视程度,以大数据的思维看待数据,也许能更敏锐地觉察读者的信息需求的变化,从而为其提供更优质的服务。

2.2.1 变样本为总体 人类曾经长期习惯于抽样、局部、片面数据,甚至仅凭经验等主观感受去探寻未知领域的规律,由于样本量有限,获得的认识难免不同程度存在着表面、肤浅、扭曲等问题。大数据时代,全数据分析模式将逐步取代样本分析模式,进而深入探寻其内在规律,获取新知。

2.2.2 变精确为混沌 “大数据”海量、多样、实时变化的特性,使人类获取数据信息的准确性降低,混沌因此而无法避免,这就意味着“小数据”时代的精确性思维向“大数据”时代的混沌性思维转变。

2.2.3 变因果为相关 “社会需要在一定程度上抛掉对因果关系的执迷,代之以简单的对应:不知道为什么,而知道是什么;这将推翻数百年的既定作法,挑战我们对于如何决策、如何理解现实的最根本的认识”^[3]。即人们只需通过相关性预测未来结果,从而

做出相关决策,而无需去探究其因果关系。例如,超市只需知道暴风雨之前,手电筒和糕点销量呈正相关,而做出将两者放置在一起销售的决定,无需去考虑其中的因果关系。

3 大数据时代高校图书馆面临的机遇与挑战

3.1 机遇

首先,大数据时代为高校图书馆数据采集提供了不竭的来源,成为新知识的主要来源。多样性特征明显的海量数据中所蕴藏的大量有价值的信息,为包括图书馆在内的各行各业拓展市场、深化服务提供了可能。以著名的“美国塔吉特(Target)超市婴儿尿布事件”为例,塔吉特公司通过数据挖掘发现某高中女生在孕婴用品页面停留较长时间,准确地预测到该女生怀孕的事实,并向其发放孕婴用品优惠券。高校图书馆服务自然理无二致,高校图书馆亦可通过对读者借阅信息的数据挖掘,精确推送其感兴趣的图书信息。

其次,大数据时代为高校图书馆提供了数据采集、分析及处理技术,如二维码技术、无线射频识别(RFID)技术等,高校图书馆应用这些技术采集、处理数据信息,深化拓展自身服务。

第三,尤为重要是“大数据”为高校图书馆提供了分析问题的思想方法。如前文所述,思维模式的变革在于——变局部为整体,变精确为混沌,变因果联系为相关性。这对高校图书馆如何敏锐发现读者信息需求的变化尤为重要。

3.2 挑战

首先,大数据时代,媒体资源愈发让人眼花缭乱,高校图书馆对读者(尤其是自控能力较弱的年轻大学生)的吸引力在下降,读者争夺将更加激烈,高校图书馆读者流失问题可能日益突出。

其次,数据分析处理服务日趋成为高校图书馆服务的主业,柜台服务将日益减少,甚至退出历史舞台,这对高校图书馆的数据处理能力、服务品质提出了更高的要求。

第三,数据安全保护难度增大,大数据技术将“数字化生存”逐渐转变为“透明化生存”,知识产权、商业利益、公共安全受损的可能性增大,高校图书馆的读者隐私安全及数据知识产权等权益将更难保障,读者隐私、知识产权被泄漏或滥用的风险在增大。

第四,大数据对高校图书馆工作人员的素质、能力提出了更高的要求,高校图书馆相关技术人才的缺乏问题将更突出。

4 大数据时代高校图书馆的对策

4.1 从非结构化数据角度突出特色资源、核心资源的建设

大数据时代,数字资源日益丰富,但也存在着馆藏资源趋于同质化的问题,因而高校图书馆从非结构化数据这一崭新的角度出发,围绕自身特色、优势学科进行专题数据采集,打造特色馆藏资源、核心馆藏资源并实时更新变得十分重要。

4.2 引进、应用与研发大数据技术

大数据技术包括海量数据的采集、存储、计算等。数据采集采用 ETL 工具 (Extraction Transformation Loading, 数据提取、转换和加载), 包括 OWB (Oracle Warehouse Builder, 适用于 Oracle 数据库的综合工具)、ODI (Oracle Data Integrator, 适用于 Oracle 数据库的数据集成类工具) 等; 数据存储包括云存储、分布式文件系统等; 数据计算包括 Hadoop 大数据分布式处理软件框架 (基于 Java 语言构建的分布式处理大数据平台)、Map Reduce (适用于大规模数据集并行运算的编程模型) 等。大数据技术的应用势在必行, 藉以提高工作效率、深化拓展服务、增强竞争力。

4.3 用数据说话, 深化拓展服务

大数据时代, 数据的重新组织、数据的再利用变得更有价值。高校图书馆积累了大量书目数据、读者数据、流通数据、数字资源利用数据、文献查询数据、参考咨询数据等等, 对这些数据进行挖掘分析, 对读者行为进行有效分析, 尤其应重视读者交互数据研究, 深入理解读者行为特征, 用科学数据说话, 这对于高校图书馆改进资源建设、深化拓展服务 (为用户提供精确、高质、高效、持续的信息资源服务, 尤其是个性化服务和精准推荐)、提高工作效率有着重要作用。

4.4 多渠道贴近用户需求

馆藏资源日趋同质化的今天, 读者资源对高校图书馆的重要性日益明显, 因而扩大高校图书馆的读者群、通过多渠道及时了解读者需求, 进而提升高校图书馆在高校教工、学生的科研工作及文化生活中的影响力日趋重要。而关注和融入微信、微博、博客、人

网、BBS (Bulletin Board System, 电子公告板) 等社交网站则是及时了解读者需求的重要途径。

4.5 切实保护读者隐私, 尊重知识产权

在“透明化生存”环境下, 高校图书馆切实保障读者隐私、尊重和保护知识产权的任务更为艰巨, 完善有关制度并严格执行显得尤为重要。

5 小结

“大数据”是高校图书馆未来发展无法逃避的发展趋势^[7], 但目前国内高校图书馆对数据的重视程度不足, 收集数据、挖掘数据意识不强, 制约了高校图书馆信息服务的进一步发展。技术上的难题往往不难解决, 思想上的轻视才是问题的关键。以信息服务工作为己任的高校图书馆, 如何以“大数据”的思维重新认识高校图书馆的信息服务工作, 如何利用大数据技术将已有的数据资源进行有序组织、深入挖掘与开发, 如何敏锐觉察读者信息需求与信息获取习惯的变化, 如何进一步推动数字图书馆、移动图书馆的发展, 成为高校图书馆未来发展的关键。

参考文献

- [1] 杨绎. 基于文献计量的“大数据”研究[J]. 图书馆杂志, 2012, 31(9): 29-32, 37.
- [2] 韩翠峰. 大数据带给图书馆的影响与挑战[J]. 图书与情报, 2012(5): 37-40.
- [3] 维克托·迈尔-舍恩伯格, 肯尼思·库克耶. 大数据时代: 生活、工作与思维的大变革[M]. 盛杨燕, 周涛, 译. 杭州: 浙江人民出版社, 2012.
- [4] 余建斌, 赵展慧. 大数据崛起[N]. 人民日报, 2013-02-22(020).
- [5] Martin Hilbert, Priscila López. The World's Technological Capacity to Store, Communicate, and Compute Information[J]. Science, 2001, 332(6025): 60-65.
- [6] 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, 50(1): 146-169.
- [7] 樊伟红, 李晨晖, 张兴旺, 等. 图书馆需要怎样的“大数据”[J]. 图书馆杂志, 2012, 31(11): 63-68, 77.

(收稿日期: 2014-03-03, 编辑: 魏民)