



零数据理论及其应用模式研究^{*}

□王彦力 杨新涯 冉蔚然

摘要 信息系统中原定设计应该产生数据的地方因各种原因不产生数据,会造成系统该部分的无效或低效。分析关注这些被忽视的地方,可及时发现系统中的漏洞与不足,从而及时调整并优化信息系统。论文提出“零数据”概念,分析零数据研究的性质及意义,构建最基本的零数据理论体系,并通过零数据理论指导图书馆行业相关工作及科研。使用零数据理论完善情报学分析体系,可以构建专门的零数据分析系统,服务于图书馆各项分析报告以及相关管理和运行机制。

关键词 零数据 零数据理论 零数据应用 阅读推广

分类号 G250

DOI 10.16603/j.issn1002-1027.2019.06.009

1 引言

信息技术的发展日新月异,移动终端迅速普及,人类生活方式随之发生巨大改变,大数据时代正式来临。近十年来,大数据已直接或间接影响了上至政府决策、金融经济、交通运输、天气预报、舆情监控等宏观层面,下至自媒体、社交、零售、休闲等与人类生存息息相关的各个领域。毫无疑问,大数据让人类走进一个新时代。但在大数据研究与应用中,人们只关注和研究了已经产生的有效数据,而许多特殊数据即本文提出的核心概念“零数据”,如系统中应该产生而没有产生的数据、没有达到设计要求的数据量以及信息系统中从未被使用的数据,在被挖掘、利用前就已经被系统清洗或淘汰了。而这些零数据对完善信息化建设、提升系统的有效性、考评运行绩效及提升服务质量等都有重要意义。因此,本文将重点研究零数据概念的内涵与外延、理论框架、产生与消除机制等基础问题,并重点探索其在图书馆行业中的应用,希望通过对图书馆领域中零数据的研究,提高图书馆行业的管理和服务效率。

2 大数据环境中的零数据

2.1 大数据研究基本概况

2011年,麦肯锡公司将大数据定义为数据量级超过传统数据库软件工具捕获、存储、管理和分析能力的数据集^[1],且并未给数据集“定义一个确定的值”。久存于军事、金融等行业的“大数据”概念因使用频率在互联网和信息行业的爆炸式增长而成为热点。多国政府对大数据表示高度重视与关注:2012年,美国政府启动“大数据研究和发展计划(Big Data R & D Initiative)”,将研究大数据上升为国家意志^[2];2015年我国国务院印发《促进大数据发展行动纲要》^[3],从国家层面部署大数据发展工作。大数据从概念的提出、发展,到如今在电网^[4]、建筑^[5]、农业^[6]等社会各行各业的广泛应用,带来了巨大的社会效益和经济价值,大数据已成为不可替代的重要资源。在学术研究方面,2008年 *Nature* 推出的大数据专刊 *Big Data* 和 2011年 *Science* 推出的数据处理专刊 *Dealing with data*,分别从“大数据处理面临的技术难点挑战”和“数据如何对社会发展产生推动作用”的方面对大数据进行了严谨探讨,在学术界引领风潮。国内自 2013 年以来,973 计划、863

* 本文系国家自然科学基金项目“智慧图书馆的零数据模型及应用研究”(项目编号:19BTQ011)的研究成果之一。

通讯作者:王彦力,ORCID: 0000-0002-8251-9089,邮箱:wangyanli@cqu.edu.cn。



计划、国家自然科学基金等都已将大数据列入重点课题,大数据成为科研工作的重点与热点。

大数据时代,数据从处理对象转化为基础性资源,同时可辅助解决其他领域的诸多问题^[7]。大数据对政府整合更多决策资源、促进提高决策成效有积极作用;为企业提供客户数据、市场发展、生产规划、上下游产业合作等各方面的支持;为学术研究开辟了几乎可以涵盖全部研究方向、且具有巨大可行性的交叉研究领域。毫无疑问,大数据及其相关技术的发展创新,将直接影响社会发展的整体趋势。

2.2 零数据在图书馆行业的出现

大数据时代,数据驱动一切发展,然而因没有产生数据,一些本应该产生数据的部分被忽略,其中存在的问题难以被及时发现,导致信息系统出现漏洞,类似的情况在图书馆行业由来已久。

如阅读推广方面,通过统计分析中国知网2006—2018年与“阅读推广”“零借阅”主题相关的学术论文发现,在论文数量上,关于“零借阅”的论文数仅占“阅读推广”论文数的2%;在被引频次上,关于“零借阅”的论文被引频次只有400余次,与“阅读推广”主题论文的总被引频次相差悬殊;在研究内容上,多是针对有数据的读者或借阅行为进行研究。但任何一家图书馆都会存在“零借阅读者”或“零借阅读籍”,如大连民族大学^[8]、广东金融学院^[9]曾专门对零借阅大学生进行调查分析,并制定对应的阅读推广方案;王静芬^[10]通过分析零借阅中文书籍,提出优化馆藏、有效选择书籍出版社及剔除复本数量等方式减少书籍零借阅现象产生。

零被引论文大量存在。研究零被引现象,对于探索科技文献的引用规律,发掘科技文献的潜在价值以及进行文献组织与管理都有非常重要的意义^[11]。零被引作为研究对象多以各杂志社、某一研究领域、情报分析为主。针对某一期刊的零被引论文进行分析综述,如《中国机械工程》^[12]、《海洋科学》^[13]等;针对综合性学科中文核心期刊的零被引论文进行分析,如以CSSCI来源期刊为范围的综合性高校学报^[14]、综合性农业科学类期刊^[15];针对某一研究领域零被引论文进行分析,如国家社科基金图书馆学领域^[16]等。这些论文,可以在一定程度上揭示零被引论文的共性,也将为科研者的选题、研究

及学术创作提供极大的参考意见。

3 零数据概念及其理论体系

3.1 零数据概念

零数据是一个尚未有明确定义的概念,在大数据时代,零数据被习惯性忽视,最终成为“不存在的数据”。而实际上零数据作为一种特殊数据,能够反映很多问题,因此唤起对零数据的关注有重要意义。

零数据是指在数据生成过程中,所收集或积累的数据集中值为0的数据。由于零数据产生的原因多样化,除了零行为不能产生数据外,一些系统误差、设备故障等,也会导致不能正常收集或遗漏收集很多数据。可见,零数据概念的外延还应包括一些与设定标准偏差极大、小到几乎可以忽略的数据,其实际意义与零数据相差无几。

3.2 零数据的性质

零数据最显著的特征是其直接性,数据显示为零,一目了然,因此相对于大数据“多样的数据类型”,零数据具有明显的单一性。相较于大数据“海量的数据规模”,零数据在信息系统中的占比非常小,出现大面积零数据的情况一般罕见,常为散状分布。但零数据一经发现就必定有问题产生,随之将会提出各种解决措施,因此零数据的价值密度较高。其改变速度与问题解决的速度成正比,经过系统优化、对症下药后往往可迅速改变;而如果不人为干预解决问题,则很难改变,故其稳定性较高。

综上,零数据具有数据显示直接、数据类型单一、数据规模相对偏小、价值密度较高、数据稳定性高的特点。

3.3 零数据的主要研究内容

目前,零数据研究是一个并不普及且容易被忽略的概念,随着对其研究的深入拓展,相关研究成果也会越来越发挥出作用和价值,亦可能成为科学研究中的“睡美人现象”。

(1) 零数据的基础理论研究

零数据的采集范围主要集中在信息系统中未产生数据的部分,研究需要时,与设定标准差异极大且数值极小的数据也纳入采集范围。零数据收集的标准则简单明了,在数据统计页面中,凡最终显示为零的数据皆为零数据采集对象,必要时小数据亦可。



零数据产生的原因是零数据理论中非常重要的部分,通过分析零数据可以直接或间接地获得不产生数据的原因,据此及时调整从而获取更多有效数据,以便更好开展下一步工作。

(2) 零数据的应用和实践研究

零数据理论与大数据理论相对相随,大数据理论开展应用的地方同样适用于零数据理论。宏观层面,安全领域中通过零数据排除安全隐患、制造业中通过零数据进行故障诊断与预测、物流行业通过零数据优化物流链并提升速度、零售业通过零数据调整产品供应与销售策略;微观层面,某视频平台可通过零数据及时下架、更新,以获取更多点击量,某商店可根据零数据调整进货出货获取更多盈利,某公司可以根据员工打卡等零数据调整工作布置、提升工作效率,某网站则可根据用户零数据及时调整服务营销策略以增长浏览量;个人层面,则可以通过零数据对自身日常生活进行检查,及时调整个人生活规划。

3.4 零数据的重要价值

在信息化日新月异的当今,数据已成为驱动发展的重要力量。零数据作为一种“特殊的数据”,其研究成果将有利于更多实际工作的开展,尤其在查漏补缺、优化系统、精细化管理等行为中不可或缺。在以下几方面有着明显积极效果:

(1) 信息系统有效性的提升。在信息管理系统建设中必然存在希望达到但实际没有达到的数据,零数据的存在可直接而迅速地反映出不足,如通过对数据的动态监测,发现数据量为零的数据集、零数据状态突然增加数据的,或者是正常数据出现异常状况,突然变成零数据等。零数据可直观反映出信息系统设定不合理、系统或活动无吸引力、设备故障、运行事故、人为干扰等不足,以便及时调整信息系统中的漏洞,优化完善系统,有效提升信息系统的有效性。

(2) 广泛应用于管理与运行绩效考评。只要用户有行为发生就必将产生数据,管理层用户数据可全面记录管理行为,客户用户数据可及时反映运行概况。在绩效考评中,这些在既定程序中产生的数据科学标准,可直接作为绩效信息,其中的零数据或小数据将成为后续工作中评定绩效以及查漏补缺的

重要参考。

(3) 开展用户行为分析。对用户数据进行分析得出目标用户的行为轨迹,从而进行精准有效的推荐。分析目标用户产生零数据的深层次原因,有助于更全面深入了解用其行为习惯,除推送其感兴趣的信息之外,更需要有针对性地采取措施,促使其尽快消除零数据,更好地使用相关信息获取更多资源及服务。此外还可根据对整体用户数据的分析,及时调整系统和产品的走向以更符合用户行为。

(4) 开展精准图书馆营销和阅读推广。对用户行为进行深入全面分析之后,用户画像的精准度将会得到极大提升,用户需求将会更加清晰明了,针对需求逐一提供其所需所急的服务,从而对用户行为进行更为贴合的精准推介。用户需求特征清晰,提供服务投其所好,服务流程清晰明了,对目标用户的宣传推广推送将明显提高效率,信息系统也将得到良好的回应。

4 零数据在图书馆行业的应用研究

随着智慧图书馆建设的不断推进,数据驱动图书馆发展已成为现实,数据的掌握与使用一定程度上决定了图书馆工作的发展与未来。零数据作为一种特殊数据在业内广泛存在,常见如零借阅馆藏纸本资源、零使用馆藏设备或资源、读者个人或团体在图书馆系统中的零数据记录、论文零被引甚至零下载、专题资源零点击量等等,是图书馆需要着重关注以不断提升服务水平的着眼点。

4.1 零数据整体应用架构

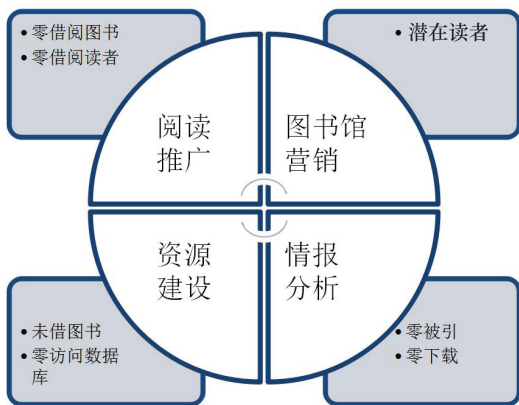


图1 零数据在图书馆中的应用架构



(1)对零数据用户的研究。在用户授权的范围,对其在图书馆馆舍及网站上的行为进行记录,会产生各项数据,统计分析这些数据,对改善服务意义重大。

零数据与大数据相比较,数据规模是微乎其微的,但通过2018年11月28日重庆大学本科生借阅量统计表(见表1)可以发现,零数据的规模(未借阅人数及其占比)还是相对比较大的。这些零数据直接反映出诸多问题,如图书馆资源配置不合理、部分读者只在馆或在线阅读但不借书、对读书活动和志愿服务不感兴趣、不愿意进入图书馆、不愿意激活用户系统、甚至根本不愿意阅读等。

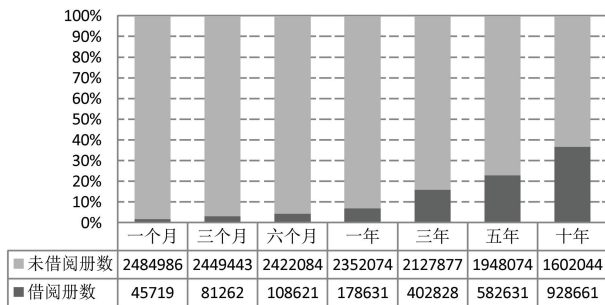
表1 重庆大学本科生借阅量统计表

年级	2018级	2017级	2016级	2015级	2014级 (已毕业)
总人数(人)	6400	6420	6408	6325	6320
未借阅人数(人)	4209	3014	1885	1229	864
未借阅人数占比(%)	65.77	46.95	29.44	19.43	13.39

对此,图书馆必须采取主动出击的方式:如针对零入馆记录读者,定期发送信息至学工系统绑定的读者邮箱、手机号,提醒读者及时入馆,直至读者产生入馆记录;针对零借阅记录读者,可以根据其检索记录,定期推送其可能感兴趣的书籍、论文等,直至产生借阅记录;针对零书评读者,可以根据其检索、借阅记录,及时推送其感兴趣的书籍,吸引其提交书评。如重庆大学图书馆在2015级新生入学两个月后通过校内邮箱给零数据读者发邮件,发出邮件后一个月内,零数据读者有明显减少。而由表1同时可以看出,重庆大学本科生人数每年基本持平,经过四年学习之后的2015级学生中,零借阅人数较之一年级新生2018级明显降低,这说明图书馆在学生生活中真实发挥了作用;而2014级学生已经毕业,但有864位学生四年都未曾产生借阅行为,非常值得图书馆深思。

(2)对零数据纸本及硬件资源的研究。借阅系统中历来存在大量零借阅书籍,2018年11月30日,对重庆大学图书馆馆藏图书借阅信息进行统计发现,未产生借阅行为的图书占据很大比例,具体如图2所示。生僻冷门、老旧残、专业性太强、小语种、品相和内容明显低劣等类别的书籍是零借阅的主

体,而这些书籍基本不会出现在阅读推广推荐的书目中,类似深圳职业技术学院举办的“从没有借过的图书”专题展览毕竟还不普及。



注:重复借阅的书籍只计算一次。

图2 重庆大学图书馆零借阅图书统计

面对这种状况,图书馆需要优化并及时调整馆藏结构,根据本校专业设置适当调整资源采购侧重点,兼顾各学科。本校未开设专业的书籍,选取经典采购,按需扩充;规模小的专业的文献,有针对性地采购;需求大的书籍尽量做好调研再采购,避免复本过多;由此避免零数据产生。老旧残书籍产生零数据的概率明显增高,应及时将其调拨入密集书库或者历史文献中心,避免占据流通率高的书籍的架位。特别需要说明的是,古籍珍本的流通率一般较低或者为零数据,因为大量古籍珍本的专业性极强且一般仅供馆内阅览,不宜突破零借阅率。

随着电脑普及率的提高,硬件借阅中的零数据也在不断增长。多所高校图书馆提供了笔记本电脑、平板电脑、kindle 借阅服务,如今借阅率明显下降,零借阅情况逐渐出现。需据此及时调整借阅模式,如增加一些便携式设备、带有专业软件的硬件设备的借阅;淘汰不必要的服务,比如可逐渐减少直至停止笔记本电脑的借阅;如条件允许,可对一些院系、团体提供有针对性的借阅,利用硬件借阅开展更多图书馆服务。

(3)对零数据数字资源的情报分析。数字资源的零数据情况最为直观,任何一个数据库都将资源的下载量、被引频次等作为重要指标进行标注,如书籍、论文等不同类型资源的零下载量、零被引频次,反映出不同类型资源的质量及水平,图书馆也会根据数据库资源的零访问、零下载、零被引等数据撰写分析与评价报告,为资源建设提供有力的数据支撑。



可见,零数据数字资源具有很强的情报分析价值。

较之“高被引”的研究热度,“零被引”的研究热度明显不足,针对该类论文进行深入分析,将对科研论文的写作提出更多有针对性的问题,为作者选题提供更多指导,规避明显的选题误区,规范写作格式等,对科研新手有非常重要的参考作用。一般而言,零被引论文或多或少存在选题偏僻、研究方向罕见、研究过时、当下已无参考意义等问题,可以很好地指导科研从业人员合理选题、适度规划、跟上时代步伐、采用详实可靠数据,少走科研弯路。

对零下载论文的研究更显冷落,多数零下载论文是由于发表时间较短导致,一定时间后都会突破零下载。但是零下载论文并非不存在,选题曲高和寡或者偏僻小众,收录期刊创刊不久或者水平低劣,论文本身价值低不值得参阅等,是零下载论文普遍存在的问题。此外也存在因访问权限设置严格而导致零下载的情况。

此外,科研专题零关注也是一个值得关注的点。各馆主页陆续设置科研专题推介栏目,旨在为科研人员提供更优质的学科服务,但如果所选专题的阅读数据为零,那么一定程度上表明大量服务被浪费。为避免零关注出现,学科馆员应主动积极与各学院、科研团体及个人沟通,深入了解科研一线人员真正所需,调动馆内资源,推出恰当的科研专题。

但特别需要注意的是,某些“零关注”的科研主题与“零被引与零下载”的论文并非完全无意义,有些研究方向虽然冷门但不可或缺、有些研究方向短期内难以产生效果但具有“科学睡美人”的潜质。因此对于零引用、零下载、零关注的科研方向,需要辩证对待。

(4)对运行过程中零数据的研究。图书馆日常工作会产生海量数据,如门禁数据、借阅数据、访问数据、下载数据等,直观地反映了日常业务工作情况。这些数据中出现的零数据揭示了工作的短板与空白,对图书馆业务的调控有着至关重要的作用。

在运行数据方面,如某分馆或专题阅览室短时间内多次出现门禁零数据,则需加大对该分馆或专题阅览室的宣传力度,或考虑是否硬件出现错误需要及时修理更换。再如馆内某运行系统出现零数据,则需尽快查清楚原因并及时调整,重庆大学图书

馆曾设计二维码占座系统,使用量一度降低为0,在分析和调研后发现主要是该馆阅览座位并不紧缺,一些没有使用系统订座的读者占据了已被预订的位置,预订该位置的同学因周边座位较多而主动选择离开,长此以往该系统的使用率渐趋于零。鉴于此,馆内将“占座系统”修改为“专座系统”,该位置一旦预订为专座就可以长期固定使用,该功能的使用频率迅速增加,零数据很快消除。

在资源建设方面,某阶段采购图书出现了大量零借阅,则需考虑加大对该部分图书的阅读推广活动,同时及时调整采购方式;某纸本资源固定时间内为零借阅率,则无需购买复本且按需调拨入历史文献中心。在用户系统方面,图书馆可根据用户的零数据进行精准推送,如每年新生必定产生大量零数据,图书馆需根据这些数据产生的学院、专业进行分学科的集中入馆培训,吸引并敦促新生使用图书馆资源;新纸本资源或数字资源启用时,根据零访问下载数据的程度,对相关专题进行侧重点不同的专题培训,引导使用;新教师入职或者新研究团队开始工作时,根据他们的零数据记录进行有针对性的学科资源、科研服务专题资源培训。总之需要根据运行记录中的零数据及时调整工作,力求完善图书馆的服务。

4.2 零数据的应用方式

(1)构建专门的零数据分析系统。目前,业内尚无专门分析零数据的系统。可构建特殊的数据分析系统,专门抓取图书馆大数据中存在的一切零数据,分析零数据产生的原因和可能性,及时提供分析报告,为图书馆的日常管理及运行提供依据。

(2)作为图书馆分析报告的重要组成部分。图书馆日常工作有大量分析报告产生,零数据应是其中重要的组成部分。在日常运行报告中,门禁、借阅、访问、下载等方面的零数据直接反映当日工作存在的问题;管理报告中,OA、文献管理中的零数据直接反映馆员工作的缺失;专题服务报告中,教学专题零关注、科研专题零关注可以表明该类专题是否有存在的必要;学科服务报告中,项目成果零被引、零下载、零收录直接反映项目的水平。可以说,零数据在分析报告中有无可替代的作用。

(3)完善相关的管理和运行机制。发现零数据



并分析其成因之后,重点是对零数据所出现的方面进行有针对性的调整和修改,优化相关措施及流程,关停明显的错误漏洞及无改造意义的部分,弥补明显缺失及错漏,强化有利因素促使其发挥更大作用,提出更多可能的解决方案。

5 结论

图书馆内零数据的产生可以直观反映管理与运行中的各种短板与缺陷,据此可迅速发现并调整相关措施,提出方案或缓解或从根源解决问题,更好为读者服务。零数据作为一种特殊的重要数据,有着巨大潜力可以挖掘,但是显然目前对零数据的重要性的认识普遍不够,人们习惯于关注已产生的数据、可能产生的数据,忽视没有产生数据的“点”,从而导致忽视零数据背后所隐藏的种种问题。关注零数据是为了通过研究零数据来及时查漏补缺,最终目的是为了消除零数据,完善图书馆的各项管理和服务工作。

参考文献

- McKinsey Global Institute. Big data: The next frontier for innovation, competition, and productivity[R]. McKinsey & Company. 2011.
- 李国杰,程学旗.大数据研究:未来科技及经济社会发展的重大战略领域——大数据的研究现状与科学思考[J].中国科学院院刊, 2012, 27(6): 647-657.
- 新华社.国务院印发《促进大数据发展行动纲要》[EB/OL]. [2018-12-25]. http://www.gov.cn/xinwen/2015-09/05/content_2925284.htm.

- 柴谦益等.基于大数据分析的智能配电网状态监测与故障处理方法研究[J].现代电子技术, 2018, 41(4): 105-108.
- 任晓宇等.建筑裂缝大数据处理平台架构[J].建筑技术, 2018(1): 61-64.
- 何婷.基于大数据处理的农业气象灾害分类模型分析[J].现代农业科技, 2018(1): 200.
- 孟小峰等.大数据管理:概念、技术与挑战[J].计算机研究与发展, 2013, 50(1): 146-168.
- 张晓丹,王学艳,陈新颜.针对高校图书馆读者零借阅现象的对策研究——大连民族大学图书馆服务创新案例[J].内蒙古科技与经济, 2017(19): 138-139.
- 黄燕.关爱零借阅大学生 点亮阅读之光——高校大学生图书馆零借阅现象调查与阅读推广对策[J].图书馆界, 2016(5): 52-55.
- 王静芬,闵翔,黄静.零借阅中文图书出版社分析与馆藏优化策略研究[J].图书馆建设, 2016(4): 58-61.
- 朱梦姣,武夷山.零被引现象:文献综述[J].情报理论与实践, 2013(8): 111-116.
- 陈勇等.《中国机械工程》近年零被引论文统计分析[J].中国机械工程杂志社, 2017, 19(4): 108-112.
- 谭雪静.《海洋科学》零被引论文分析[J].编辑学报, 2017, 29(1): 93-95.
- 张崧颖.综合性高校学报零被引论文分析——以CSSCI来源期刊为例[J].出版广角, 2017(23): 50-52.
- 杨晓容等.综合性农业科学类中文核心期刊零被引论文特征[J].中国科技期刊研究, 2017, 28(5): 468-473.
- 程竹仪.国家社科基金图情档领域的零被引论文分析[J].卷宗, 2018(2): 25-26.

作者单位:重庆大学图书馆,重庆,400044

收稿日期:2018年12月1日

(责任编辑:支娟)

Research on Zero Data Theory and Its Application Model

Wang Yanli Yang Xinya Ran Weiran

Abstract: The original design of the information system was supposed to generate data. The failure of doing so due to various reasons however results in ineffectiveness or low efficiency of this part of the system. Analyzing and paying attention to these areas can help find the system vulnerabilities and deficiencies immediately, so as to adjust and optimize the information system in time. This paper proposes the concept of “zero data”, analyzes the nature and significance of zero data research, constructs the zero data concept system and guides the related work and scientific research in the library industry through the theory of “zero data”. The zero data concept system is used to guide the library and information science, constructing a special data analysis system to serve library analysis reports and related management as well as operation mechanism.

Keywords: Zero Data; Zero Data Theory; Zero Data Application Model