

· 业务研究 ·

基于用户数据资源化利用的高校图书馆服务创新研究

张群¹ 李爱国²

(1. 江南大学图书馆, 江苏 无锡 214122; 2. 东南大学图书馆, 江苏 南京 210096)

〔摘要〕 [目的] 研究用户数据资源化利用, 以促进高校图书馆服务创新。[方法] 归纳分析了图书馆4种不同类型的用户数据, 介绍了用户数据的采集方法, 并对大数据环境下图书馆用户数据的特征进行了分析。在此基础上, 提出了基于用户数据资源化利用的高校图书馆服务创新架构和模式: 建设用户信息数据库, 提供精准个性化服务; 搭建即时交互式平台, 提升用户体验服务; 构建泛在知识环境, 尝试泛在图书馆服务; 适应用户需求转变, 提供知识化服务。[结果/结论] 高校图书馆应高度重视用户数据的资源化利用, 创新图书馆服务的内容和形式, 促进图书馆事业新发展。

〔关键词〕 用户数据; 资源化利用; 服务创新; 高校图书馆

DOI: 10.3969/j.issn.1008-0821.2017.05.018

(中图分类号) G252 (文献标识码) A (文章编号) 1008-0821 (2017) 05-0111-05

Study on the University Libraries' Service Innovation Based on User Data Resource Utilization

Zhang Qun¹ Li Aiguo²

(1. Library, Jiangnan University, Wuxi 214122, China;

2. Library, Southeast University, Nanjing 210096, China)

〔Abstract〕 [Objective] This paper studied user data resource utilization to promote the university libraries' service innovation. [Method] It made an inductive analysis of four different types of library user data, introduced user data collection methods, then analyzes characteristics of library user data under the big data environment. Based on the user data resource utilization, it proposed university libraries' service innovation architecture and pattern: construction of user information database to provide accurate personalized service; establishment of Real-time interactive platform to improve user experience service; Building ubiquitous knowledge environment to try ubiquitous service; adapt to meet the needs of users to provide knowledge service. [Result/Conclusion] University libraries should pay high attention to the utilization of user data, so as to innovate the university libraries' service content and form and promote the new development of library cause.

〔Key words〕 user data; resource utilization; service innovation; university library

随着科学研究的第四范式——数据密集型科研时代的来临, 科学数据量呈几何级增长。Brooks Hanson (2011) 提出了“数据驱动科学发展, 科学就是数据, 数据就是科学”^[1], 表明了数据对于科学发展的重要性。大数据蕴藏着无穷的价值, 将促使各行各业的革命性改变。用户是图书馆的服务对象, 是图书馆创新发展的源泉和动力。图书馆作为高校信息中心和信息服务提供者, 应积极地把握大数据带来的机遇, 加强对用户数据的资源化和增值利用,

促进服务内容纵深发展, 最终实现信息共享和知识创新。

1 图书馆用户数据归纳分析

用户数据是指用户基本信息及其在使用图书馆的过程中产生及关联的各种信息。图书馆可获得的用户数据按数据类型不同可分为用户基本信息、需求信息、行为信息、反馈信息四大类。对这海量、无序的用户数据进行资源化

收稿日期: 2017-01-31

基金项目: 国家社会科学基金一般项目“大数据时代图书馆用户信息的资源化研究”(项目编号: 13BTQ025)。

作者简介: 张群(1976-), 女, 副研究馆员, 硕士, 研究方向: 图书馆知识服务。

利用,分析用户的行为特征,掌握用户的数据需求特征,中可获得的数据具体归纳分析如表1所示。能够全面提升图书馆的服务水平。图书馆与用户交互过程

表1 图书馆用户数据归纳分析

序号	数据类型	数据内容	数据来源
1	基本信息	统计学信息	年龄、性别、单位、家庭地址、毕业院校、学历、电邮、电话、个人照片等
		关联信息	学习经历及成绩、论文、专著、教材、专利、项目、课件、设计图、试验产品、获奖情况等
2	需求信息	数据采集	数据管理计划的制定、数据库选择、检索词及检索策略确定等
		数据生产	实验设计、调研方式和方法的选择、元数据等
		数据存储和管理	数据管理计划的实施、数据储存设施的确定、数据质量筛选、数据评价等
		数据保存和共享	数据的上传、共享、查询、下载、权限管理、用户管理等
		数据引用和出版	数据挖掘、数据发现、数据引用、知识产权和版权、隐私和安全问题等
		流通借阅	借阅文献种类、借阅册数等
3	行为信息	参考咨询	咨询方式、咨询内容、咨询时间等
		数据库利用	检索、浏览、下载、复制等
		科技查新	委托合同、背景技术、查新目的、查新报告等
		文献传递	文献内容、传递方式、文献格式等
		培训讲座	内容、方式、时间、人员等
4	反馈信息	反馈内容 用户互动 反馈方式	网站留言、MOOC 互动、微信点赞、用户反馈信息表、用户评价记录

2 图书馆用户数据的采集

表1中图书馆用户数据的数据按照结构分,包括3类:一类是各类数据库中记录的用户基本信息、文献借阅信息等结构化数据,数据规范性强,结构清晰,使用方便;一类是用户在利用数据库时生成的日志记录等半结构化数据,通常为HTML页面、XML文档形式,数据量大,价值密度低,需要进行抽取利用;另一类是用户反馈信息、行为信息等非结构化数据,如文档、图片、报表等等。对于这3类数据有不同的采集方法:①结构化数据的采集:结构化数据存储于关系数据库中,具有一定逻辑结构,可直接通过接口进行调用,使用方便;②半结构化数据的采集:半结构化数据具有结构,但结构变化大,能实现非结构化数据与结构化数据结构自由转换,需要根据用户IP或者事先设定好的用户标识,提取用户的日志信息,并设定规则对数据进行清洗,提取保存用户的使用行为记录;③非结构化数据的采集:非结构化数据没有固定结构,不能将其标准化,不能保存于关系数据库,只能以文件形式存放,所以采集、存储以及使用这些非结构化数据需要智能化的系

统,该系统能够实现非结构化到结构化数据转换,提供统一的仿真接口完成各个异构数据源的融合,实现数据的一致性,消除数据的异构性^[2]。从异构数据源融合集成用户数据,关键在于不同系统中用户标识要具有惟一性^[3]。

3 大数据环境下图书馆用户数据的特征分析

大数据,即数量巨大、结构复杂、类型众多的数据构成的数据集。在大数据环境下,图书馆用户数据具有了与以往截然不同的崭新的特点。对大数据环境下图书馆用户数据的特征进行分析,有助于将多元化的用户数据序化为用户数据资源,并开发利用用户数据资源、提高用户数据资源价值。

3.1 数据量更大且分布更广

图书馆可获得的用户数据不仅包括用户基本信息和行为信息,也包括科研、教务、档案等管理系统中关联的用户信息,还包括用户评价、反馈等信息,数据量更大。同时,数据分布性更广,主要体现在图书馆用户数据会分布、驻留在图书馆各种应用与服务系统。如:用户基本信息存在于用户管理系统中;用户需求信息和行为信息遍布于

OPAC系统、查新委托平台等图书馆应用管理系统中; 用户利用数据库产生的日志保存于不同数据库商的服务系统中; 用户反馈信息存在于网站留言、微博、微信等系统中。另外, 用户置身于泛在网络环境, 其信息行为突破了时空限制, 可以在任何时间、地点, 利用各种移动终端使用图书馆; 可以方便地借助网络技术以电子邮件、Web、BBS、讨论组、QQ、微博、微信等方式与图书馆进行互动。这就使得用户数据呈现出分布性、发散性的特点。并且, 图书馆用户数据来源渠道也更多: 移动互联网、物联网、手机、电脑、传感器等, 都可以成为图书馆用户数据的来源。

3.2 数据异构化程度更高

大数据环境下, 数据形式更加灵活多样, 如数据库数据、文本、文档、表格、音频、视频文件等。数据的存储形式不同, 数据所在的服务器不同, 操作系统的开发平台不同, 诸多的差异使得数据之间无法互换、共享。根据数据格式的不同, 数据大致可分为结构化数据、半结构化数据和非结构化数据, 但只有15%左右的信息有效存储于结构化数据库^[4]。数据异构化, 使得用户数据资源化的难度加大。应用Web服务技术、SOA技术、XML技术等新技术, 将非结构化数据转换为结构化数据, 存入关系数据库中统一管理, 实现异构数据的转换、融合, 将是图书馆用户数据资源化的研究重点。

3.3 数据的动态性更强

图书馆用户数据是图书馆与用户在交互活动中产生并逐步积累起来的。其中用户基本信息中的统计学信息是相对静态的, 而用户需求信息、行为信息和反馈信息是动态的, 呈现出多样化、动态性、多变性的特点。尤其对于移动用户来说, 移动终端设备、网络环境、行为情景的不同, 使得移动用户的信息需求与传统用户具有更显著的区别特征, 用户移动搜索的目的更为多样化, 如检索学习辅助资料、浏览时事新闻、社交沟通、娱乐休闲、位置导航等。日志是最重要的动态数据, 能够反映用户的个性化特征, 是图书馆确定核心用户、分析用户需求、预测用户信息行为等的重要依据。通过图书馆日志的知识发现研究, 可以了解用户的使用方法、搜索特征、数据库的访问率等, 从而改进图书馆主页建设、资源采购、业务流程等。

3.4 数据的复杂性更高

图书馆的用户数据从产生的机制来看, 有动态与静态之分; 从数据格式看, 有结构化、半结构化数据和非结构化之别; 从来源来看, 有人工录入和自动获取之异^[3]。另外, 传统的图书馆用户信息行为中, 用户只是图书馆信息资源的利用者, 与图书馆进行单线互动, 用户之间基本上没有互动, 而在社会网络大数据环境下, 用户可以参与各种社会软件或者社会网络站点(如Facebook、YouTube、Wikipedia, 人人网、豆瓣网等), 通过交流和互动获得多

方信息, 集信息利用者、创建者和传递者于一体。社会网络是一种全新的信息交流模式, 具有强烈的参与性和互动性, 吸引了用户的热点关注和广泛参与, 已成为用户在网络世界交流的主阵地。为了让社会网络中的用户能方便地利用图书馆, 同时也吸引潜在用户, 国内外高校图书馆已开始尝试图书馆与社会网络站点的融合, 如在Facebook、YouTube、人人网、豆瓣网上的图书馆。这种模式不仅增大了图书馆的信息空间, 也使得图书馆用户数据更为复杂多样。

3.5 信息安全形势更加严峻

国际图联规定: 图书馆用户享有个人的信息隐私权, 图书馆员及其所属机构应保护用户的信息隐私, 不论其使用何种形式的信息、享受何种形式的服务^[5]。大数据时代, 人们通过智能搜索以及强大的机器运算能力, 可以对不同来源的数据进行关联、对比、分析, 从而锁定个人的真实身份, 个人的隐私和尊严将不可避免地受到伤害。同时, 基于网络服务的数字图书馆系统, 拥有分布广域性、体系结构开放性、信息资源共享性和信道共用性等特点, 导致系统难免脆弱^[6]。这些使得用户信息安全形势更加严峻。图书馆如何采用身份验证、加密、访问控制等多种技术的集成来保证用户信息安全, 是用户数据资源化建设首先需要考虑的问题。

4 基于用户数据资源化利用的高校图书馆服务创新

图书馆零散、无序的用户数据大多分散储存在图书馆各个系统中, 没有进行整合, 也没有深入开展数据分析和挖掘工作, 不能作为资源被图书馆或其它服务系统所利用。用户数据资源化利用是从海量、分散、有噪声、异构的用户数据中提炼出潜在的、有价值的用户知识、模型或规则, 并与其他知识生产、组织和服务系统开放关联, 以利于图书馆科学运行和服务创新。

基于用户数据资源化利用的高校图书馆服务创新架构如图1。

具体来说, 基于用户数据资源化利用的高校图书馆服务创新主要包括以下4种模式:

4.1 建设用户信息数据库, 提供精准个性化服务

用户信息是图书馆提供精准的个性化服务的前提。图书馆应通过全方位、多渠道地采集用户信息, 如通过数字资源管理分析数据库进行数字资源用量分析、资源分布、热点统计、下载监控等, 结合用户基本信息、需求信息等, 完美地抽象出用户的信息全貌, 全面构建用户信息数据库, 并应用智能数据挖掘软件系统, 精准分析用户需求信息, 实现精准用户服务。如, 通过对用户性别、职业、借阅记录等统计数据的挖掘, 预估用户的潜在需求; 通过纵向分析用户的信息行为, 掌握用户信息行为的变化规律; 通过分析用户的反馈信息, 了解用户对图书馆的满意度、忠诚

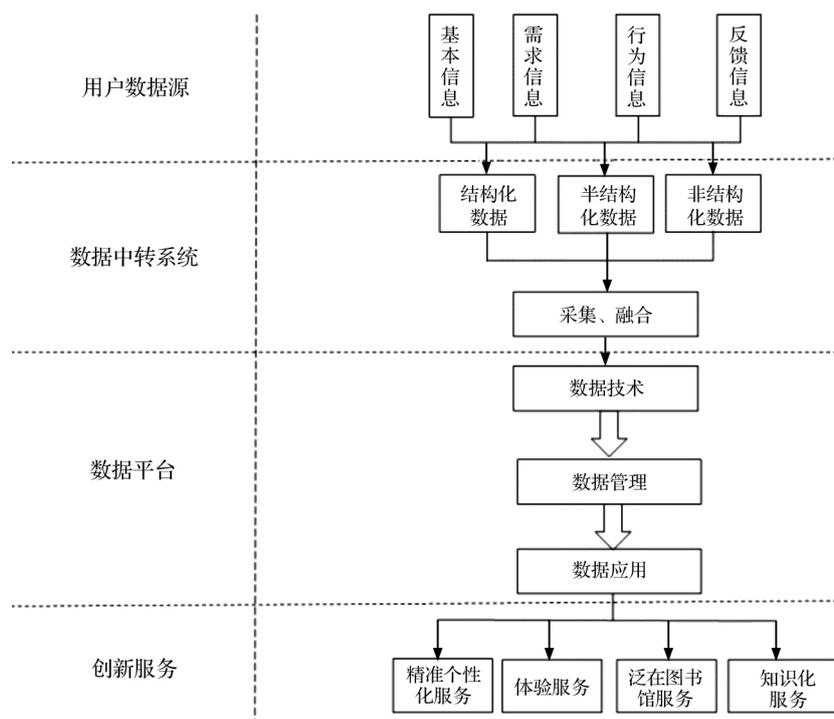


图1 基于用户数据资源化利用的高校图书馆服务创新架构

度，针对不同目标用户群体实施精准的个性化服务。

4.2 搭建即时交互式平台，提升用户体验服务

用户不仅是图书馆服务的对象，同时也是信息的传递者、生产者、评价者。为提高用户的忠诚度和满意度，首先图书馆应建立以用户为中心、具有交互功能的一站式信息检索平台，使用户能方便快捷地获得信息；同时采用“超链接”技术，在用户需求的信息资源旁进行相关内容的链接，实现“知识发现”功能；其次，借助微信、微博、QQ等新媒体，加强图书馆与用户之间的信息交互，建立用户参与服务创新的激励机制和评估机制，促进用户体验、评论、讨论、分享信息；第三，图书馆可以尝试与社会网络站点的融合，以扩大图书馆的用户群，也扩展图书馆在网络空间的影响力，如，耶鲁大学科学图书馆网页集成了社交网站 Facebook 和 MySpace；清华大学图书馆在人人网平台上建立了书友会并分享了“人人网校内——图书搜索”^[7]。

4.3 构建泛在知识环境，尝试泛在图书馆服务

泛在知识环境是指由网络设施、硬件、软件、信息资源、人等有机组成的新一代科技知识基础结构，通过计算、存储和通信方面的最大便利，使资源能够更为彻底地发挥作用而构建的一种普遍的、综合性的知识环境^[8]。泛在知识环境下，高校图书馆的服务将以智能化、简单化、人性化为特征，以交流、协作为中心并基于通用技术使资源最大限度的重复使用^[9]。泛在图书馆服务立足用户需求，倡导“用户在哪里，服务就在哪里”，为用户提供一种无处

不在、无时不有的服务。目前图书馆能提供的泛在服务模式主要包括：①移动图书馆：利用手机、PDA等移动终端作为接受、传播服务的主要渠道，解决了时空障碍；②嵌入式服务：嵌入式学科服务和嵌入式桌面信息服务将图书馆服务置于用户的信息环境，即时高效、无缝对接；特别是一些高水平大学实行的“学科分馆——学科馆员”模式，将使学科服务在创新模式中得以长足发展^[10]；③图书馆云服务：指基于云计算的各项服务。“云计算”将计算任务分布在大量计算机构成的资源池上，提供了高度虚拟化的服务，如 SaaS 服务，使用户真正做到“按需即用、即需即用”；④基于 MOOC 的在线服务模式：图书馆通过多种方式参与 MOOC，如参与开放存取问题的讨论，进行“信息素质教育”、“数据素养教育”等课程设计和制作，或将图书馆员“嵌入”MOOC 环境从而向学习者提供信息支持。

4.4 适应用户需求转变，提供知识化服务

用户能够方便地利用搜索引擎、网络数据库等获得大量的公开信息，因此，用户所需求的信息也不再是单一的、常规性的、事实型的信息，而是经过加工的二次、三次信息或者具有“知识特征”的综合性信息，图书馆以馆藏文献的流通阅览和简单咨询为传统的信息服务方式急需向以数据揭示、数据管理、数据挖掘、数据关联、数据发现等为主的数据知识化服务模式转变。比如：①建立整合用户信息数据库、专家库、机构知识库、特色资源库、学科导航库等为一体的科学数据知识服务平台，并为科学研究

提供科学数据管理规划制定、数据分析、数据处理、数据库管理等方面的技术支持服务; ②以科技查新带动学科服务及技术竞争情报服务的开展, 如产业导航、企业运营类专利导航等; ③以定题和专题情报服务带动数据分析、数据挖掘等服务, 为用户提供决策支持; ④以机构知识库、特色资源库、学科导航库等的制作促进数据资源开发与利用; ⑤挖掘知识内容, 构造可扩展和可视化的知识地图, 形成支持创新的知识空间; ⑥应用信息智能技术服务促进数字参考咨询服务, 如本体学习技术、语义网技术、推送技术等。

在数据密集型环境下, 高校图书馆应高度重视用户数据的资源化利用, 通过多渠道采集、融合用户数据, 采用数据挖掘、数据关联等技术, 掌握用户数据特征, 构建用户画像, 有助于图书馆迅速定位用户群体, 优化调整图书馆运行策略, 创新图书馆服务的内容和形式, 实现高校图书馆从信息中心到科学数据中心的转型, 促进图书馆事业新发展。

参 考 文 献

[1] Brooks Hanson, Andrew Sugden, Bruce Alberts. Making data

(上接第46页)

[17] Frischbier S, Petrov I. Aspects of Data-Intensive Cloud Computing [J]. DBLP, 2010, (6462): 57-77.
[18] Hu W, Yang T. The Good, the Bad and the Ugly of Consumer Cloud Storage [J]. Acm Sigops Operating Systems Review, 2010, 44 (3): 110-115.
[19] Horrigan J. Use of Cloud Computing Applications and Services [EB/OL]. <http://pewinternet.org/Reports/2008/Use-of-Cloud-Computing-Applications-and-Services.aspx>, 2017-01-19.
[20] MacAskill E. WikiLeaks Website Pulled by Amazon After US Political Pressure [EB/OL]. <http://www.guardian.co.uk/media/2010/dec/01/wikileaks-website-cables-servers-amazon>, 2017-01-19.
[21] Amazon Blocks WikiLeaks but Denies 'Pressure' [N]. China Daily, 2010-12-04, (8).
[22] Peter G, Noriaki I. Hybrid Cloud Management: Foundations and Strategies [J]. Review of Business and Finance Studies, 2013, (4): 37-50.
[23] Chakraborty R, Ramireddy S, Raghu T. The Information Assurance Practices of Cloud Computing Vendors [J]. IT Professional, 2009, 12 (4): 29-37.
[24] Kim DY, Park ST. A Study on Effects of Security Risks on Acceptance of Enterprise Cloud Service: Moderating of Employment and Non-employment using PLS Multiple Group Analysis [J]. Journal of Computer Virology and Hacking Techniques, 2016, 5 (3): 1-11.
[25] Caragnano G, Goga K, Ruiu P. Scalability of a Parallel Application in Hybrid Cloud [R]. Complex, Eighth international conference on intelligent and software intensive systems (CISIS), 2014:

maximally available [J]. Science, 2011, 331 (6018): 649.

[2] 万里鹏. 非结构化到结构化数据转换的研究与实现 [D]. 成都: 西南交通大学, 2013.
[3] 李爱国, 汪社教, 孟祥保. 图书馆用户信息模型框架构建 [J]. 图书情报工作, 2015, (13): 13-19.
[4] 张德政, 张萍萍. 非结构化信息管理 [J]. 微计算机信息, 2006, (9): 218-219.
[5] 汪强. 读者个人信息保护有法可依 [J]. 图书馆工作与研究, 2014, (4): 8-12.
[6] 王碧琴, 任洁, 冯彦平, 等. 数字图书馆用户信息隐私的安全威胁分析 [J]. 图书馆学研究, 2015, (10): 34-36.
[7] 孙彩杰, 庄小峰. 图书馆融入社会网络的案例分析 [J]. 图书馆建设, 2010, (11): 57-60.
[8] 武群辉, 李卫峰. 基于用户需求的图书馆泛在化服务模式 [J]. 图书馆学刊, 2012, (7): 95-97.
[9] 郭瑞芳. 泛在知识环境下高校图书馆信息资源的构建 [J]. 情报资料工作, 2010, (4): 55-58.
[10] 沈洋. 泛在知识环境下的高校学科服务模式构建研究 [J]. 现代情报, 2017, (2): 75-77, 111.

(本文责任编辑: 孙国雷)

451-456.

[26] Mann A. The Cost Benefit Myth of the Public Cloud [EB/OL]. <http://pleasediscuss.com/andimann/20110504/the-cost-benefit-myth-of-the-public-cloud>, 2017-01-19.
[27] Julisch K, Hall M. Security and Control in the Cloud [J]. Information Security Journal, 2010, 19 (6): 299-309.
[28] Khan KM, Malluhi Q. Establishing trust in cloud computing [J]. IT Professional, 2010, 12 (5): 20-27.
[29] Anthes G. Security in the Cloud: Cloud Computing Offers Many Advantages, but Also Involves Security Risks [J]. Communications of ACM, 2010, 53 (11): 16-31.
[30] Hon W. The Problem of 'Personal Data' in Cloud Computing: what Information is Regulated? The Cloud of Unknowing [J]. International Data Privacy Law, 2011, 1 (4): 211-228.
[31] Khan AN, Kiah MLM, Ali M. A Cloud-Manager-Based Re-Encryption Scheme for Mobile Users in Cloud Environment: a Hybrid Approach [J]. Journal of Grid Computing, 2015, 13 (4): 1-25.
[32] 彭海深. 云计算技术浅析 [J]. 微计算机信息, 2010, (10): 176-178.
[33] 孔德新. 私有云下的身份与管理解决方案 [J]. 信息安全与通信保密, 2012, (11): 75-77.
[34] Smith v. Maryland [EB/OL]. <http://laws.findlaw.com/us/442/735.html>, 2017-01-19.
[35] Soghoian C. Caught in The Cloud: Privacy, Encryption, and Government Back Doors in The Web2.0 Era [J]. Journal on Telecommunications & High Tech. Law, 2009, 8 (2): 359-371.

(本文责任编辑: 孙国雷)