

数据驱动下的高校图书馆服务创新 ——以华东师范大学图书馆为例

汪志莉, 于亚秀, 程煊
(华东师范大学图书馆, 上海 200062)

摘要: 探讨数据驱动下高校图书馆服务创新, 给出从数据层至应用层的数据资源整合利用框架图, 总结数据驱动下高校图书馆服务的新特点, 即个性化、可视化、集成化、共享云开放平台、主动推动。并举例基于数据的个性化服务、决策支持、接口开发、开放获取等方面的实践, 希望能为高校图书馆数据服务提供新的思路, 以应对当前高校图书馆数据服务面临的挑战, 促进高校图书馆数据服务进一步发展。

关键词: 个性化; 可视化; 集成化; 共享云开放平台; 主动推送

中图分类号: G254

DOI: 10.3772/j.issn.1673-2286.2016.8.011

在移动互联网和大数据的强烈冲击下, 传统的阅读方式、资源组织方式以及图书馆服务方式都受到巨大冲击和挑战, 图书馆开始思考其在学校教学、科研等各类活动中的角色和定位, 并大胆尝试运用各种新的手段和技术推出新型服务, 使其融入科研、教学、管理过程。国外图书馆在数据驱动的读者服务方面有很多成功的应用案例, 例如, VIVO建立的科学家网络^[1], 康奈尔大学面向科研的资源发现服务等^[2]; 国内现阶段还处于理论与实践起步阶段, 例如, 清华大学推出的文献计量分析服务^[3]、华中科技大学开展的科技情报服务等^[4]。国内外利用数据驱动服务创新的案例为本研究的开展提供借鉴。

如清华大学邓景康馆长所说“虽然图书馆的数据规模以及对数据处理的实效性远未达到大数据处理的需求, 但是基于大数据开展的挖掘数据价值、提取知识的理念却深深影响了图书馆的服务模式”^[5]。结合实践案例, 探讨高校图书馆数据驱动下的服务创新, 给出从数据层至应用层的资源整合利用框架, 以读者为中心, 努力寻求服务与数据的契合, 最大可能释放图书馆基础数据的价值, 提高资源利用率, 在基于数据的个性化服务、决策支持、接口开发、开放获取等方面

进行尝试, 希望能够促进高校图书馆数据服务的进一步发展。

1 高校图书馆数据资源整合利用框架

图书馆数据资源整合利用框架图(见图1), 分为三层, 第一层为数据层, 介绍高校图书馆主要数据来源, 包含两大部分, 资源数据和用户数据, 资源数据包含4大类, 如表1所示, 数据通过描述规范进行定义, 通过平台互操作标准以及各类链接标准进行交换和共享, 用户数据包含两方面, 如表2所示, 一般通过各系统的统计报表或者日志进行查询和分析; 第二层介绍图书馆数据服务的新特点, 即个性化、可视化、集成化、共享云开放平台、主动推动(该部分在文章第二部分重点阐述); 第三层介绍各类数据消费接口, 即数据如何与图书馆、校园网内其他应用进行数据交换与共享(文章第三部分举例基于数据的个性化服务、决策支持、接口开发、开放获取等方面的实践)。

这些数据环境及应用提醒图书馆人员, 图书馆有条件也急需进行服务创新, 应该突破馆藏的边界, 利用全网域的资源 and 数据, 为用户提供更多、更好的集成服

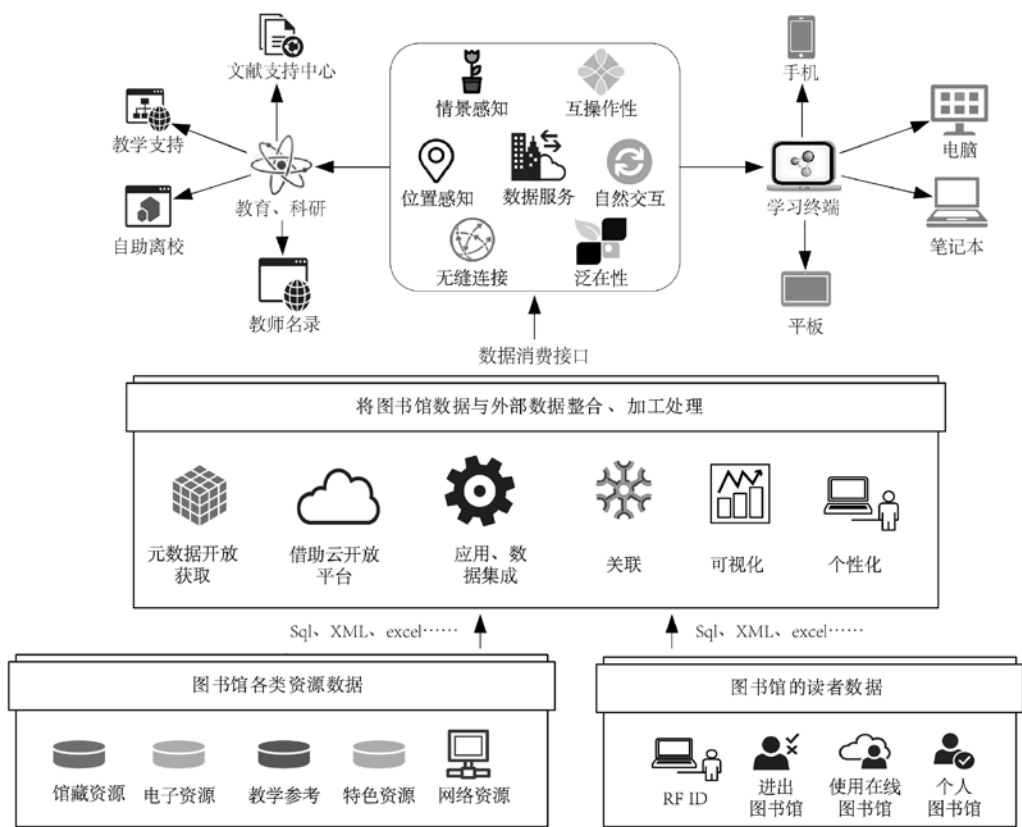


图 1 图书馆数据资源整合利用框架

表 1 资源数据

分 类	示 例	数据描述规范	平台的互操作标准	链接标准
实体馆藏资源	馆藏书目数据	XML、MARC21、CNMARC、RDA	Z39.50	-
特色数字化资源	硕博学位论文库、方志库等	XML、CNMARC、都柏林核心	-	-
订购的电子资源	中国知网、万方数据库等	-	OAI-PMH Web Services	DOI OpenURL
互联网资源	wiki词条、优酷免费音视频等	-	-	DOI OpenURL

表 2 用户数据

分 类	示 例
到馆产生的数据	进馆通道机数据、借阅量、借阅历史、研讨间使用等
在线行为数据	主页访问、自助系统、数据库访问、数据下载等

务，满足学校教学、科研各类活动的数据支持，应该努力寻求服务与数据最大可能的契合点，最大可能释放图书馆基础数据的价值，扩大图书馆服务范围、提高图书馆服务质量、改善用户使用体验。

2 数据驱动下图书馆服务的新特点

图书馆服务，必须和大学图书馆的使命相匹配，顺应时代发展。就目前国内图书馆的现状，总结为三点^[6]：

第一, 要体现服务性, 服务于教、学、科研活动, 服务于教师、学生、职员、社会读者, 服务于大学建设目标, 服务于知识生产、保存、传播和利用; 第二, 要具有层次性, 坚守基础服务 (例如馆藏建设、全媒体馆藏揭示、全网域馆藏发现等), 努力探索服务提升模式 (例如提供更加精确、个性化的服务, 提供知识发现、开放获取等等); 第三, 要具有差异性, 为各类读者提供量身定制的服务, 尽可能做到个性、专业、全面, 令每位读者满意。

2.1 个性化

用户信息需求日趋个性化, 而高校图书馆传统的信息服务模式难以满足用户的这种需求。个性化信息服务已成为高校图书馆深化信息服务的重要内容^[7]。国外在20世纪90年代末就开始开发和设计图书馆个性化服务系统, 美国于1997年开始投入使用MyLibrary系统, 其后英、法、日、德等国也开始这方面的研究^[8]; 2010年谢菲尔德大学联合OCLC, 开发基于用户需求与偏好的图书馆目录推荐系统^[9]; 相较于国外, 国内图书馆个性化服务开始较晚, 2001年我国开始图书馆个性化信息服务的研究和应用, 中国科学院国家科学数字图书馆根据北卡罗来那州立大学的MyLibrary系统成功开发了“我的数字图书馆”; 国内图书馆近几年在推荐系统中取得部分成果, 如上海图书馆在“MyLibrary”系统中增加畅销书推荐的功能^[9], 目前大部分高校都有“我的图书馆”栏目, 但与国外相比, 其对信息挖掘分析的深度和精确度仍不够, 这将是我国目前及下一个阶段研究的重点和难点。

2.2 可视化

数据挖掘已经研究多年, 目前很多学校都已将自身的基础数据通过程序处理加工, 形成报告或者文档, 但是基于文本的服务模式已经远不能满足现代读者对信息摄取时限的要求。这就需要更简单、直接、可视化的方式来展示数据, 推送数据服务。目前比较流行的是用数据说故事的方式, 即将众多数据进行融合、归纳、对比, 通过生动鲜艳的视图方式展示, 以增强视觉冲击, 使读者在尽可能短的时间内摄取尽可能多的信息。在设计上, 可以借鉴比较成熟的商业化产品的展示方式, 如百度迁徙、景区热力图、百度慧眼、通勤图等, 直观、清晰、一目了然。虽然还有差距, 但是图书馆数据

服务已经在往这个方向努力。譬如, 早期OPAC书皮、图书推荐、OPAC上显示图书书架等; 近几年推出的毕业季图书馆记忆, 如厦门大学的“囿·时光”^[10]、华东师范大学的“校园记忆之图书馆生活”^[11]、上海交通大学的“《印·迹》纪念册”^[12]等; 还有目前比较热门的关联数据研究, 如2016年上海图书馆上线的家谱知识服务平台^[13], 其查询系统可根据人物、地点进行地图可视化查询, 是基于关联数据可视化展现的试点, 这些都说明图书馆正不断创新数据服务呈现方式, 以应对互联网发展的需求。

2.3 集成化

如何利用全网域资源提供更加集成化的服务, 是图书馆目前研究的热点和趋势。集成包含应用层面的集成和数据层的集成, 目前大部分图书馆都已实现应用层的集成, 如OPAC的资源集成, 百度直达号、易企秀等轻应用的服务集成等; 数据层的集成目前主要是利用关联数据技术实现。国外在资源数字化集成方面, 多为实践或成果, 有代表性的案例概述如下。2008年, 欧洲多媒体在线图书馆 (Europeana) 项目正式对公众开放^[14], 它是一个集合欧洲数字资源的门户网站和搜索引擎, 其元数据采用RDF存储, 目的是方便在语义环境中通过关联数据对其进行有效揭示, 提高资源可用性; 2013年, 美国数字公共图书馆 (Digital Public Library of America, DPLA) 项目正式对公众开放^[15]; DPLA将美国档案馆、图书馆、博物馆和文化遗产机构、私人收藏机构等分散的资源进行统一集合, 旨在探索如何建设一个开放、分布式的在线资源网络; 2011年DPLA宣布与Europeana合作, 建立可互操作的数字模型、资源规范、可开放获取的馆藏资源并共享源代码; Europeana、DPLA及DPLA与Europeana合作的实践项目均有基金会支持; 2009年, OCLC利用SRU服务为虚拟国际规范文档 (the Virtual International Authority File, VIAF) 项目提供关联数据^[16], 这些数据来自14家图书馆使用不同格式表示的16个权威文档, 目标是通过匹配和关联不同语言国家图书馆的规范文档, 以提高利用率。国内基于关联数据技术在资源数字化集成应用方面, 文献较少, 图书馆界对语义网技术的研究以基础概念介绍辨析、本体创建、关联数据发布流程与技术、书目数据的语义化居多。有代表性的有欧石燕提出面向关联数据的语义数字图书馆资源描述与组织框

架^[17]、刘炜^[18]、夏翠娟等^[19]基于W3C的RDB2RDF标准规范基础上,对关联数据的应用进行探索和实践,典型案例是基于BIBFRAME建立上海图书馆的家谱本体,并发布为关联数据。利用关联数据进行大规模异构平台的数据集成,目前刚起步。

2.4 共享云开放平台

大数据和互联网时代,对图书馆馆员自身信息素养和计算机编程能力都提出更高的要求,特别是图书馆系统部和情报咨询部门,要求馆员能用计算机编程实现数据加工和批量处理,实现用程序代替人工机械重复、繁杂的工作。但是论团队和专业程度,图书馆数据服务和百度、谷歌、超星等知名公司的差距是显而易见的。所以馆员应尽可能借助云开放平台激活图书馆服务,灵活修改应用代码,即可将成熟的产品直接应用于图书馆的各类服务,不仅节省人力开发成本,而且产品的稳定性、推广性、可移植性都更优越。

2.5 主动推送

改变传统图书馆被动服务的模式,对基础数据再组织再加工,并关联各系统数据,形成个性化、可定制的数字资源。利用互联网和大数据技术,将资源和服务主动推送给电脑、移动用户,使读者突破馆舍的束缚,随时随地享受图书馆的服务。对图书馆的资源,尽可能实现开放获取,对有争议的知识产权,可通过开放元数据的方式推送数据,并针对不同的服务对象提供不同的数据和资源,最大限度挖掘图书馆的资源,提高资源利用率。2013年,华东师范大学图书馆将特色资源硕士、博士论文数据库的元数据推送给研究生院、教师名录,通过平台开放实现元数据的开放获取,改传统被动获取为主动推送,为学位论文数据库增加附加值和新的访问点,提高数据资源的利用率。

3 数据驱动的创新服务及实施效果

3.1 个性化的服务——“个性化图书推荐活动”

鉴于目前高校图书馆OPAC上“我的图书馆”栏目缺乏对信息的深度挖掘与关联,2015年华东师范大学

图书馆推出基于用户日志的个性化图书推荐活动,通过挖掘借阅历史数据,分析读者兴趣喜好,满足读者个性化服务的需求,提升纸本图书借阅量。

图书推荐系统基于华东师范大学图书馆集成系统(Integrated Library System)中的图书外借日志,利用ApacheMahout库编程进行推荐生成,通过sed、awk、shell脚本等工具将推荐转换为数据模型所支持的偏好数据格式,选择合适的推荐算法组件,通过编程生成图书推荐记录,借助Z39.50协议匹配图书ISBN号从而获取图书完整信息^[20]。读者登录OPAC个人借阅账户后,浏览器中运行jQuery代码以读者ID为参数发送异步请求,接收到请求的服务器端程序查询推荐记录数据库,并以JSON格式返回推荐记录,jQuery代码以图片链接的形式动态呈现推荐图书。

3.2 为图书馆管理层提供决策支持——“馆藏文献利用调查与分析”

高校图书馆目前普遍存在借阅率逐年下降的现象^[21],鉴于此,华东师范大学图书馆于2009年起对馆藏文献进行系统地调查分析,希望通过客观数据指导图书馆馆藏发展、藏书布局、流通借阅规则设置、服务人员配置等,以期为读者提供更好的服务,满足读者需求。

华东师范大学图书馆2009年首次推出《可借阅纸质文献利用情况调查》(2010年改名为《馆藏文献利用调查与分析》报告^[22]),框架主要包含五大方面(馆藏基本情况、图书馆利用情况、借阅情况、借阅文献分析、馆藏与借阅的关系分析)和附件的“借阅排行”。框架的设计始终将各种指标与读者捆绑在一起,并对读者群按照各自的特点进行划分,使数据更能反映客观事实。通过数据分析,对全面掌握读者对印刷型资源的借阅行为、了解网络环境下读者对实体图书馆的需求变化、引导读者阅读、优化图书馆服务、合理调配资源等方面都起到积极促进作用。

3.3 为职能部门提供决策支持服务——“利用H指数对教师科研论著影响力进行测评”

高校图书馆始终以向学校教学、科研提供文献信息支撑服务为己任,高校管理却很少被纳入图书馆信息服务范畴,事实上,高校管理层也需要大量信息为其管理决策提供支撑服务^[23]。尤其在当今争创全球知名高

水平研究型大学的背景下,我国高校想要取得跨越式发展,教学、科研、管理、保障服务缺一不可。其中,管理决策对学校的发展又起着至关重要的作用。图书馆应发挥管理文献资源的优势,开展基于文献资源的揭示、发现及获取服务,并对图书馆常规服务进行延伸,寻找学校需求与图书馆可提供数据的契合点,努力为学校管理提供决策支持服务。

3.4 元数据开放获取——“利用学位论文资源开展社会服务的实践”

学位论文作为高等学校重要的特色资源之一,反映一个学校图书馆最具特色和学术性的收藏,是学生学术成果的提炼,也是教师从事教学及科学研究的重要成果^[24]。将图书馆收藏的学位论文数据库与校内其他系统进行整合,为其增加附加值和新的访问点,扩展学位论文数据库作为特色馆藏的使用范围和价值,以增加大学与社会、企业合作、交流的机会。

华东师范大学图书馆通过学位论文回溯建库工作,实现了较为完整的数据收录。通过开发基于平台和数据库的学位论文元数据接口开放元数据,并将元数据嵌入学校其他应用或网站,实现利用学术资源开展有针对性的社会服务。针对学位论文数据库系统本身的特点,采用WCF平台进行面向应用的Open API开发,提供学位论文信息服务Web Service,让不同系统的第三方应用根据需要调用学位论文元数据Open API,实现同构或异构平台间的数据共享和互操作。同时采用轻量级的Web Widget技术,将小巧的客户端学位论文信息服务调用代码嵌入用户浏览的网页中,实现学位论文信息的直接推送。

该项工作的实施提高了学术资源元数据可见度,推进学术资源元数据开放,扩大馆藏特色学术资源服务范围,为高校图书馆开展社会服务开辟新思路。

4 结论

本文对数据驱动下图书馆的服务创新进行探讨,提出在大数据及互联网时代图书馆应该拓宽思路,尽可能挖掘图书馆潜在数据资源,扩大图书馆服务的读者群体,使图书馆服务与学校的教学、科研相融合,为读者提供有价值的数据参考,引导读者使用图书馆各类资源,提升图书馆各类资源的使用率;对图书馆常规

服务进行延伸,有针对性地为教学、职能部门提供定制化数据服务,探讨图书馆开辟决策支持服务的新途径;通过将学位论文元数据嵌入第三方平台,达到推进学位论文资源的开放与共享,提升学位论文作为学术资源的社会服务功能,释放图书馆基础数据价值的目的。

在实施过程中也存在许多困难和挑战,譬如个性化图书推荐活动的使用率及推荐成功率并不高。分析原因主要有两方面,其一,数据本身受到邱均平等研究印证的数据稀疏性条件下推荐效率客观情况的影响^[25];其二,宣传力度不足,为此可借助新媒体工具,如微信、微博、百度直达号、易企秀等,使读者及时了解图书馆推出的新服务。此外,还需要及时进行用户体验调查,调整算法、改善界面,使服务从内容到形式都更符合读者需求。而利用学位论文资源开展社会服务的实践,由于受到知识产权等限制,只能进行元数据层面有限的揭示与共享。希望今后图书馆能加大数据集成、开放共享,图书馆员能不断学习,为图书馆的资源组织管理、服务质量提升和服务范围拓宽等,提供更加行之有效的技术支持。

参考文献

- [1] VIVO:Enabling the national networking of scientists[EB/OL].[2010-12-02].
<http://www.vivoweb.org>.
- [2] Cornell University Library.Research and expertise across Cornell.
[EB/OL].[2016-03-05].<http://vivo.library.cornell.edu/>.
- [3] 清华大学图书馆.文献计量分析服务[EB/OL].(2015-08-17)[2016-03-05].
<http://lib.tsinghua.edu.cn/service/docqua.html>.
- [4] 华中科技大学.科技情报服务[EB/OL].[2016-03-05].<http://www.lib.hust.edu.cn/ArticleChannel.aspx?ChannelID=214>.
- [5] 邓景康.大数据环境下清华大学图书馆的实践[EB/OL].(2013-08-29)
[2016-08-01].http://www.tsinghua.edu.cn/publish/news/4207/2013/20130829152841109507827/20130829152841109507827_.html.
- [6] 龚天芳.漫谈“图书馆角色定位”与实践[EB/OL].[2015-12-18].<http://gaoxiao.cnki.net/2015guangzhou/list-pdf-1801.html>.
- [7] 夏南强,张红梅.基于数据挖掘的数字图书馆个性化服务[J].图书馆学研究,2006(1):32-34,43.
- [8] 阮光册,夏磊.推荐系统的发展与公共图书馆个性化信息服务探讨[J].图书馆,2016(2):94-99.
- [9] 谢菲尔德大学和OCLC获得开发图书馆推荐系统的资助[J].现代图书情报技术,2010(9):83.
- [10] 13th IFLA BibLibre international library marketing award winners

- announced[EB/OL].(2016-02-18)[2016-08-01].<http://www.ifla.org/node/10258>.
- [11] 华东师范大学图书馆.2015年毕业生:校园记忆之图书馆生活[EB/OL].(2015-06-20)[2016-08-01].<http://202.120.82.52:8082/index.asp?p=10865301>.
- [12] 上海交通大学图书馆《印·迹》纪念册[EB/OL].(2015-06-15)[2016-08-01].<http://memory.lib.sjtu.edu.cn>.
- [13] 上海图书馆家谱知识服务平台[EB/OL].(2016-03-05).<http://jp.library.sh.cn/jp/home/index>.
- [14] Europeana collections. Explore 53 657 626 artworks, artefacts, books, videos and sounds from across Europe [EB/OL].(2016-01-18).<http://www.Europeana.eu>.
- [15] DPLA. Primary source sets:fresh perspectives on history,culture,and literature [EB/OL](2016-01-20).<http://dp.la/>.
- [16] 王薇,欧石燕.关联数据在图书馆领域的应用研究[J].新世纪图书馆,2012(9):25-28.
- [17] 欧石燕.面向关联数据的语义数字图书馆资源描述与组织框架设计与实现[J].中国图书馆学报,2012(6):58-71.
- [18] 刘炜.关联数据:概念、技术及应用展望[J].大学图书馆学报,2011,29(2):5-12.
- [19] 夏翠娟,刘炜,张磊,等.基于书目框架(BIBFRAME)的家谱本体设计[J].图书馆论坛,2014(11):5-19.
- [20] 刘丹.利用Apache Mahout部署个性化图书推荐服务[J].现代图书情报技术,2015(10):102-108.
- [21] 王秀芳,钱小荣,李静.高校图书馆期刊馆藏发展现状与策略研究[J].情报杂志,2010(12):174-175.
- [22] 2014年度华东师范大学图书馆馆藏文献利用调查与分析[EB/OL].(2016-08-01).http://www.lib.ecnu.edu.cn/about/lib_annual/.
- [23] 宋振世,周健,吴士蓉.h指数科研评价实践中的应用研究[J].图书情报工作,2013(1):117-121,135.
- [24] 李欣,姜爱蓉.大学图书馆开展社会服务再思考[J].图书馆工作与研究,2013(12):55-58.
- [25] 邱均平,张聪.高校图书馆馆藏资源协同推荐系统研究[J].图书情报工作,2013,57(22): 132-137.

作者简介

汪志莉,女,1983年生,硕士,馆员,E-mail: zlwang@library.ecnu.edu.cn。
于亚秀,女,1986年生,硕士,馆员。
程煊,女,1981年生,硕士,馆员。

The Data-driven Innovation of Service of University Library: a Case Study of ECNU Library

WANG ZhiLi, YU YaXiu, CHENG Xuan
(Library of East China Normal University, Shanghai 200062, China)

Abstract: Confronted with challenges from university library data service, this article has discussed how to innovate university library service driven by data, with an aim to provide new thoughts for university library data service and promote its further development. It first described the integration diagram of data resources from the data level to the application level and then summarized its new features, namely individualized, visualized, integrated, sharing cloud open platform and message pushing. In addition, it has given examples concerning applications of individualized services, decision-making, portal development and open access.

Keywords: Personalizing; Visualization; Integration; Shared Cloud Open Platform; Active Push

(收稿日期: 2016-07-18)