



2018年 CASHL 前瞻性课题研究项目

图书馆数字人文服务研究及实践 (数字人文服务专题信息平台及培 训工作坊建设)

课题单位：华东师范大学图书馆

课题负责人：于亚秀

课题组成员：周健（华东师范大学图书馆）、宋振世（华东师范大学图书馆）、陈枝清（华东师范大学图书馆）、程静（华东师范大学图书馆）、鲁丹（华东师范大学图书馆）、张毅（华东师范大学图书馆）

结项时间：二零一九年五月

摘要：图书馆为学校教学科研提供专门的、专业性的优质服务是图书馆的重要任务之一。数字人文研究的活跃，对图书馆的人文服务提出了新的要求，人文研究者普遍存在数字技术欠缺的短板，不仅体现在不会使用技术，还体现在不会用数字人文的思维方式思考人文问题，相当多的图书馆正积极思考如何为数字人文研究提供服务支持。

课题组在开展图书馆数字人文服务研究是实践过程中，本着既不能偏离图书馆的主要职能，也不能脱离数字人文的客观性需求的原则，高校图书馆承担着文献收集整理、开展咨询服务以支持本校教学研究需要的根本职能，因此，我们在开展数字人文服务的研究及实践种主要围绕图书馆在资源建设、信息素养咨询培训等隶属于图书馆本职服务等方面展开的。

关键词：数字人文服务 数字人文工具 数字人文工作坊 资源优化 信息素养培训

目录

1 课题背景	4
2 前期调研及项目团队	4
2.1 国内外调研	4
2.2 项目团队	6
3 华东师范大学图书馆数字人文服务实践	6
3.1 资源优化——以 CASHL 特藏《杜鲁门口述历史全集》为例	6
3.1.1 资源介绍	6
3.1.2 资源格式优化方法及工具	7
3.2 数字人文专题信息平台建设	10
3.2.1 资源内容	10
3.2.2 平台建设	10
3.3 数字人文工作坊	11
3.3.1 数字人文信息素养培训	11
3.3.2 数字人文沙龙	12
4 结语	12
参考文献	13

1 课题背景

数字人文是人文研究在信息技术高度发达的今天，充分发挥计算机技术的数据处理能力、传播能力，对传统研究方法和研究范式的改革和创新。图书馆在信息时代也正经历着由初级的以藏书为主，向着充分利用现有信息资源，通过信息技术、管理制度的有效开发和设计，为各类科学研究提供专门的、专业性的优质服务转变。数字人文的发展也给图书馆的发展带来了新的思路和契机^[1,2]。相当多的图书馆正积极思考如何为数字人文建设提供支持，充分利用现有图书馆资源开展数字人文服务。为学校教学科研提供专门的、专业性的优质服务是图书馆的重要任务之一，数字人文的活跃，对图书馆的人文服务提出了新的要求。

图书馆在开展数字人文服务的过程中，既不能偏离图书馆的主要职能，也不能脱离数字人文的客观性需求，高校图书馆承担着文献收集整理、开展咨询服务以支持本校教学研究需要的根本职能，因此，我们开展数字人文服务主要围绕图书馆在资源建设、信息素养咨询培训等隶属于图书馆本职服务等方面展开的。

在 2018 年 CASHL 前瞻性课题研究项目的支持下，本研究的实践工作从以下三个切入：一是优化馆藏资源，以 CASHL 大型特藏之一的《杜鲁门口述历史全集》缩微平片为研究对象，通过优化资源格式，方便读者做一进步的文本分析和文本挖掘，拓展资源的使用价值。二是调研、整理、汇聚数字人文前沿信息和工具，建设专业的、有针对性的数字人文信息专题平台，并建立平台内容可持续性更新机制；三是调研国内外数字人文培训工作坊课程及模式，以及校内不同读者群体的实际需求，探索并开展针对人文研究的不同过程、不同读者群体的，有针对性的、高质量的培训和讨论会。积极探索图书馆以主动服务和被动服务相结合、多渠道的模式开展数字人文信服务实践的方式。

2 前期调研及项目团队

2.1 国内外调研

(1) 美国情况

“数字人文”是一个在近二十年内在人文科学领域形成的子领域，在欧美国家已羽翼丰满，早在 2006 年美国国家人文基金就成立了“数字人文创新项目”，后更名为“数字人文办公室”。CenterNet 是国际数字人文中心网站^[3]，CenterNet 上的数字人文中心统计数据能真实反映当前国内外数字人文中心建设现状。通过检索、地址位置标注查阅核实，在美国的 73 个人文中心中，由学校内部组织的数字人文中心为 57 个，占到 72.6%。学校是数字人文中心建设的主要单位。我们进一步对美国数字人文中心的主办单位进行分析^[4]，发现主要由学校、院系、图书馆、联盟或协会等。分析数据如下：

① 图书馆是美国数字人文中心建设的重要力量

在校内，由学校和院系主办或合办的数字人文中心为 40 个，占 55%，由图书馆主办的数字人文中心为 12 个，加上图书馆和学校其他院系一起联合主办的有 5 个（图 1），即在

隶属于学校层面且由图书馆主办或参与主办的数字人文中心占 23%。

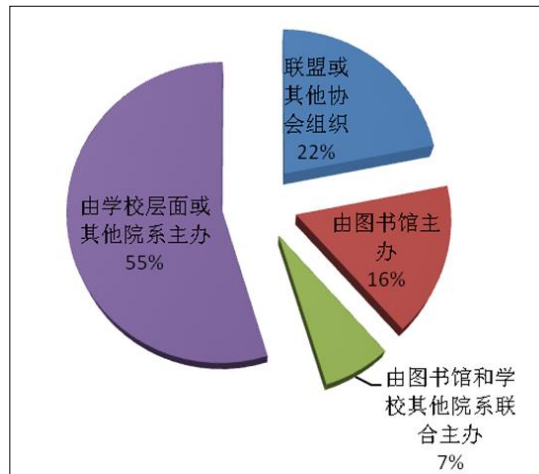


图 1 CenterNet 登记的数字人文中心图书馆参与情况

② 图书馆是美国数字人文中心建设的重要支柱

虽然 57 个数字人文中心中，有 40 个数字人文中心非图书参与主办。但我们通过对其人员组成进行调研发现，在 27 个可详细查阅人员组成的数字人文中心网站上，其中 11 个人文中心的组成成员有图书馆员参与。其他未能查阅人员组成的数字人文中心网站，也能在其主页介绍或者项目介绍的文字中明确发现与图书馆相关的描述。如：非图书馆参与主办的弗吉尼亚大学的数字人文中心（The Institute for Advanced Technology in Humanities），尽管成员组成也没有图书馆员，但在其项目内容介绍中对图书馆的合作和参与进行了感谢。

由以上两个层面的分析数据可以得出，美国高校图书馆在数字人文服务方面有很高的参与度，图书馆也是数字人文研究实践的重要力量。

（2）国内情况

在国内，港澳台地区的数字人文起步相对较早，1998 年台湾就发起了“数位博物馆计划”，并从 2009 年连续开始举办“数位典藏与数位人文国际研讨会”。大陆数字人文起步相对较晚，但近几年，国内图书馆开始重视数字人文理论支撑下开展相关活动。继 2011 年武汉大学成立国内首家数字人文研究中心之后，2016 年和 2017 年北京大学和南京大学分别成立数字人文小组；2014 年上海图书馆发起了“数字人文与语义技术”学术会议，并于 2016 年 2 月推出基于关联开放数据的数字人文服务——上海图书馆家谱知识服务平台；2015 年开始多个内地高校积极参与数字人文研究，如北京大学、清华大学、南京大学、南开大学、上海大学等，纷纷举办了与数字人文相关的主题会议，通过会议凝聚共识，促进发展。例如：北京大学主办的“数字人文论坛”，已经举办三届，主题分别是“跨界与融合：全球视野下的数字人文”^[5]、“互动与共生：数字人文与史学研究”^[6]，和“孵化与实践：需求驱动下的数字人文项目”^[7]，国内数字人文活动也逐步从理论研究落地到实践。此外，2017 年 4 月北京大学图书馆举办首期“数字人文工作坊”系列活动，也开启了国内图书馆落地实践数字人文服务的又一项重要活动。

虽然国内外图书馆参与数字人文中心建设都能够充分发挥图书馆的馆藏资源丰富、数字化加工技术熟练、用户群体庞大、访问便捷等优势，并与其肩负的信息咨询和文化传播、保存和传承历史文化等社会功能目标相一致。但是国内图书馆对数字人文学科关注不够，许多建设和应用实践仍处于自为状态，缺乏数字人文理论支撑下的资源建设，不利于革命性的升级服务模式，服务于数字人文环境下的人文研究范式，还带来资源建设的重复和浪费。所以以面向数字人文教学研究为目标，借鉴美国在数字人文领域建设的先进经验，发挥图书馆资源优势，全面开展图书馆参与数字人文服务的实践研究，对于促进我国数字人文研究具有积极意义，也能深入挖掘图书馆在研究领域的贡献潜力。

2.2 项目团队

合理的团队分工很重要，往往涉及多个部门参与，有些工作可同时进行，比如资源 ORC 加工、资源整理、平台设计等，在图书馆人力紧缺的情况下可缩短建设周期。

资源建设方面：由数字化建设部负责资源的优化处理，由平台与技术支持部负责资源阅读方式的优化。

数字人文信息平台建设方面：由平台与技术支持部开整理数字人文工具、研究支持部整理数字人文相关项目、书籍、论文，平台与技术支持部开发平台并将整理的信息进行展示。

数字人文工作坊建设方面：图书馆内部开展数字人文沙龙，以及数字人文信息素养培训。同时与华东师范大学调查与数据中心联合面向全校举办系列数字人文讲座。

3 华东师范大学图书馆数字人文服务实践

3.1 资源优化——以 CASHL 特藏《杜鲁门口述历史全集》为例

数字人文服务对资源的要求，已经需要图书馆从资源的被动提供者向信息资源的加工者转变。图书馆资源是数字人文研究者在研究活动中寻找第一手资料时重要的考虑对象，对于数字人文研究者来说，收集和准备所需的材料是一项耗时的工作，学者们需要花大量时间寻找、组装和准备原始资料。图书馆可以考虑充分优化馆藏资源，对资源进行元数据开发、多源数据的融合和语料库构建，使数字人文学者能够在图书馆获得完备的、可直接用于人文研究的第一手资料。

以图书馆数字化自建资源为例，只是将纸质文献扫描为 pdf 的电子文献，其文件格式不能很好地适应读者的科研需求。将 pdf 文件文本化，可以为读者提供更适合研究的资源呈现方式和服务方式，方便读者做一进步的文本分析和文本挖掘，拓展资源的使用价值。

在本次项目中，我们以本馆自建数据库《杜鲁门口述历史全集》为第一着手点，优化其资源形态，方便其用于人文研究。

3.1.1 资源介绍

《Harry S. Truman Oral Histories Collection Microfiche Complete with Guide (杜鲁门口述历史全集)》是华东师范大学历史系/周边国家研究院梁志教授极力推荐的有

关回忆杜鲁门总统生活和时代事件的一套缩微资料，共 628 张缩微平片，由 ProQuest 公司出版，2015 年 CASHL 专项经费引进，华东师大图书馆是该文献国内高校唯一收藏馆。这套文献集中关注了杜鲁门政府时期美国对外政策以及援助计划的制定者，还原了这些外交界和军界重要人物对柏林空运、联合国、遏制共产主义以及中国大陆与台湾冲突等问题的历史记忆与观察。此外，还采访了不少杜鲁门政府时期负责内政的官员，追寻当时他们对“公平施政”经济计划、劳工骚乱、当代民权运动起源以及其他重要问题的认知。在数字化前，该文献资料还未能从其它数据库中查阅到，因而具有很高的史料价值。

该套缩微平片在 2017 年 CASHL “特藏++” 深度服务项目的支持下，对其中的 396 种实现数字化，由缩微胶片转换为 pdf 并在网站上展示，但资源的格式并不利用人文研究者进行文本分析，校外用户也只能通过口述者、口述内容摘要等简单元数据了解文件内容，不利于校外人文学者判断是否需要文献传递等共享服务。

3.1.2 资源格式优化方法及工具

在项目期间，我们对已数字化的 396 种《杜鲁门口述史全集》资源进行资源优化，完成了以下工作（图 1）。

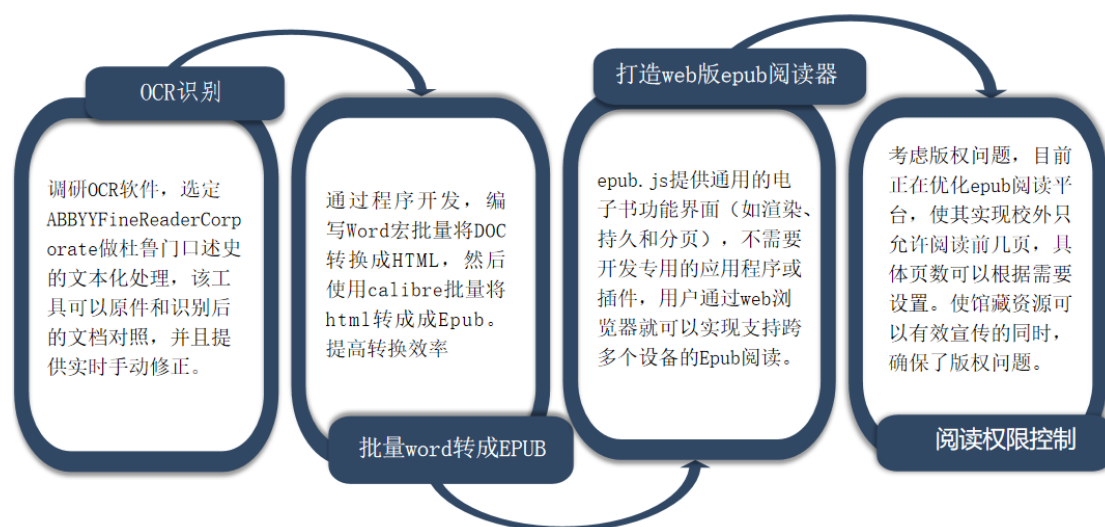


图 1 《杜鲁门口述史全集》资源优化流程

(1) OCR 识别

BBYY FineReader 是专业的 OCR 文字识别软件，在全世界名列第一，他能够提供精准和高效的文档识别、数据提取解决方案，支持多个国家的字符和彩色文件识别，主要用于将图片型 PDF 转化成可编辑的文本，他还拥有强大的编辑和转换功能。

(2) 批量 word 转成 epub

将 word 转换成 epub 的方式很多，为了提高转工作效率，我们首先通过程序开发，编写 word 宏批量将 OCR 识别后的 .htm 转换成 .epub 文件（图 2），然后通过 Calibre 批量将 .htm 转换成 .epub（图 3），极大的提高了工作效率，OCR 加工后的 396 种资料在半小时内便可以完成 word 到 epub 的转换。

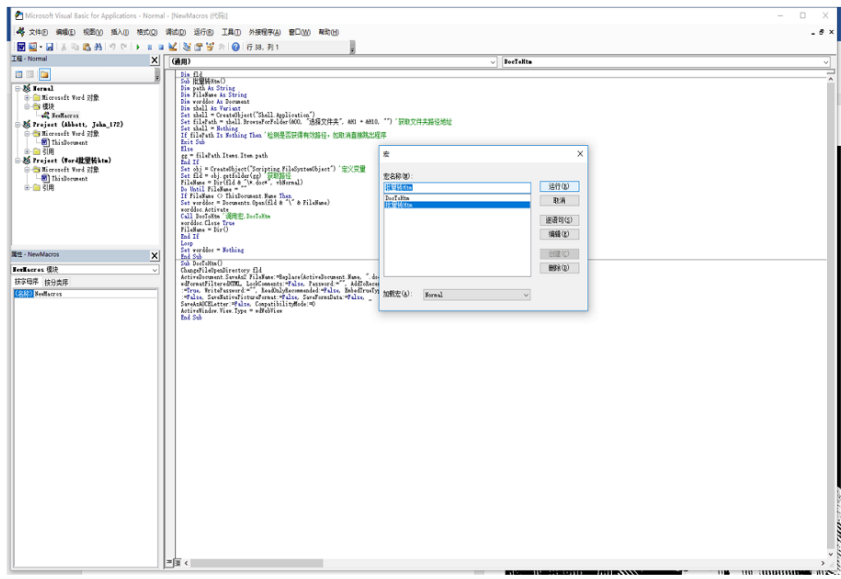


图 2 word 宏将 .doc 文件批量转换成 .htm

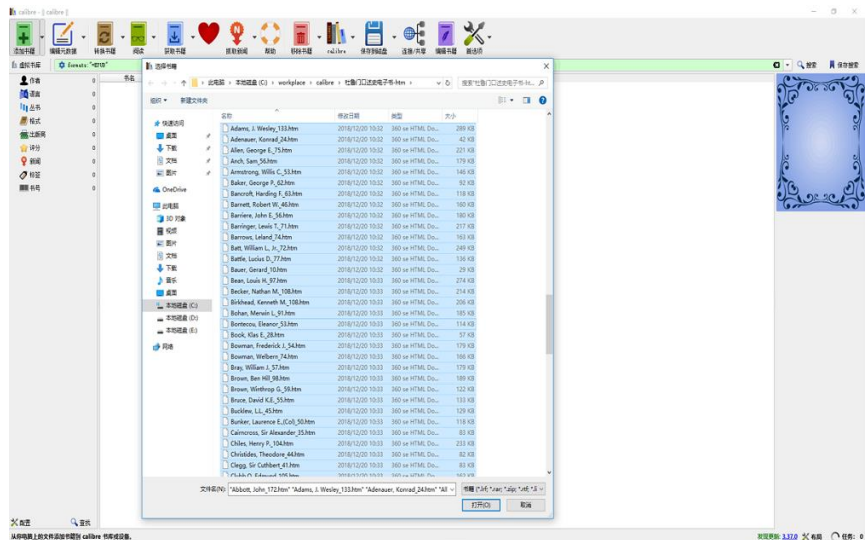


图 3 Calibre 批量将 .htm 转换成 .epub 文件

(3) 打造 web 版 epub 阅读平台

EPUB 标准是一种广泛使用的和易转换格式。许多书是目前这样的格式，它可以转换为其他格式（如 PDF、Mobi 和 iBooks）。一个解压缩的 EPUB3（你可以将 epub 文件用解压缩软件进行解压）是集 HTML5 文件、CSS、图像和其他媒体，并且允许我们根据受控词汇呈现书的部分内容和一整本书。

传统的 epub 阅读器需要下载并本地安装，给读者带了很多不便，我们通过 epub.js 进行二次开发，打造 web 版 epub 阅读平台，epub.js 是一个在浏览器端渲染 epub 文件的 JavaScript 库，支持在跨设备的浏览器中呈现 epub 文档，提供通用的电子书功能界面（如渲染、持久和分页），不需要开发专用的应用程序或插件。开发完成的 epub 阅读平台，只需要将 epub 格式电子书上传到共享服务器：\\202.120.82.**，即可以在 PC 电脑或者移动

设备浏览器中阅读该书，书的访问链接 http://202.120.82.**/epub/?书名。（图4，图5）

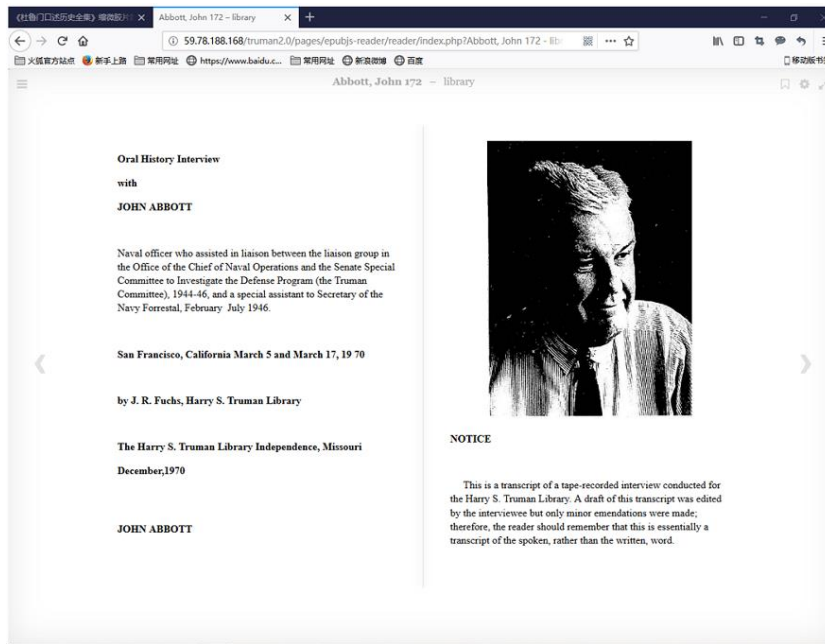


图4 web版epub阅读平台PC端阅读界面

(4) 优化《杜鲁门口述史全集》缩微平片数据库平台

在以上工作的基础上，我们进一步优化2017年完成的《杜鲁门口述史全集》缩微平片数据库平台，考虑阅读权限问题，校内ip地址范围内浏览增加epub阅读功能（图5），校外ip浏览增加epub试读功能，使其实现校外只允许阅读前几页，具体页数可以根据需要设置。使馆藏资源可以有效宣传的同时，确保了版权问题。（图6）。



图5 校内epub阅读功能

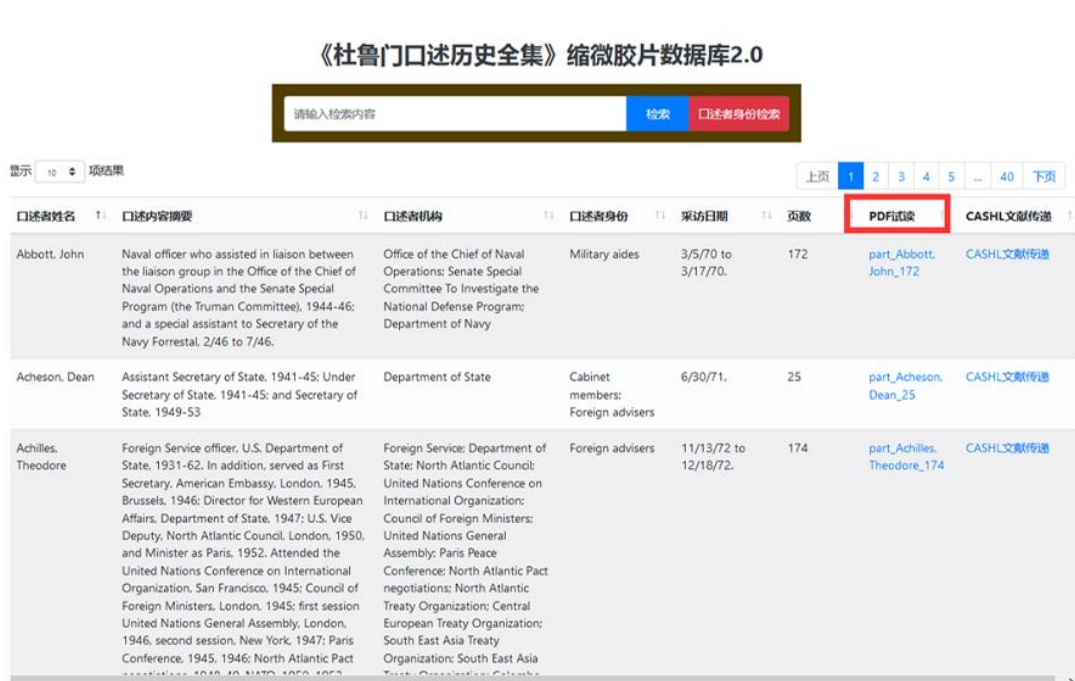


图 6 校外 epub 试读功能

3.2 数字人文专题信息平台建设

3.2.1 资源内容

专题信息平台是人文学者从图书馆获取帮助的重要渠道，其不受地理空间限制的优势，可以最大化数量的为人文学者提供帮助。华东师范大学通过整理数字人文学科介绍、国内外数字人文相关出版物、国内外数字人文中心及链接、数字人文工具（按开源/商业工具进行分类、按功能类型分类）、国内外数字人文社区（协会、联盟、会议等）等资料，搭建数字人文学习平台，为数字人文学习者提供第一手学习资料。该平台的完善是持续的过程，后续我们将进一步完善工具的整理，及教程的上传。

3.2.2 平台建设

我们通过开源软件 wordpress 搭建本校的数字人文学习平台（图 7），WordPress 是使用 PHP 语言开发的 blogging 平台，用户可以在支持 PHP 和 MySQL 数据库的服务器上架设属于自己的网站，功能强大、扩展性强，是快速搭建网站的工具之一。

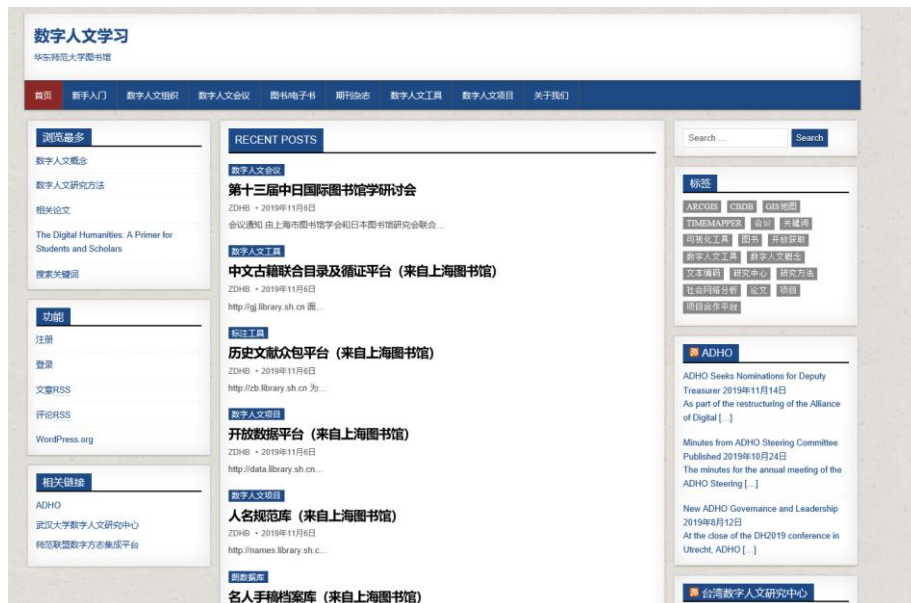


图 7 华东师范大学图书馆数字人文学习平台

建设汇聚整理归纳来自于互联网的数字人文前沿信息和工具的专题信息平台并不困难，搭建专题信息平台也是图书馆可以快速介入到数字人文服务的方式。为了更好更快的开展数字人文服务，为人文学者提供快速获取数字人文信息的渠道，华东师大正在依托网站检索收集整理资料、数字人文微信群里发布的相关信息，整合构建数字人文工具汇总平台。与此同时，为了更好的、可持续性的提高该平台的质量，华东师大也在努力开展专职数字人文馆员岗位角色设定和深化培训。

3.3 数字人文工作坊

数字人文文科和理工科的领域交叉领域，技术欠缺是人文研究者的短板，在国外很多高校也在都已经建立了综合考虑课程设置、课程内容、培训对象等多个方面的健全数字人文培训模式^[8]，不仅帮助读者快读的提高数字人文素养、掌握一定的数字人文工具，也提升了图书馆在人文研究者中的地位、拉近人文学者和图书馆人文服务的距离。

华东师大图书馆在数字人文信息素养培训方面也紧随数字人文学科的发展，在信息素养教育课程中增加了相关课程，并从以下两个方面展开。

3.3.1 数字人文信息素养培训

华东师范大学图书馆与面向全校师生开展讲座：华东师范大学调查与数据中心与图书馆联合，面向全校师生，举办系列数字人文信息素养讲座。（图 8，图 9）

中北校区

数据库 工具软件 数字人文		讲座地点：中北校区图书馆一楼数字阅览室B201	
序号	讲座名称	日期	课件下载
1	青研联动注意力数据建设与分析系统	2019-2-27 13:00-14:30	下载
2	Data Citation Index——揭开尘封的宝藏	2019-5-2 13:00-14:30	下载
3	Derwent Data Analyzer——文献大数据分析神器	2019-5-8 13:00-14:30	下载
4	MCA toolbox——文献数据多维分析及可视化软件	2019-5-15 13:00-14:30	下载

图8 华东师范大学中北校区图书馆系列数字人文讲座

数据库 工具软件 数字人文		讲座地点：闵行校区图书馆一楼读者培训室B107	
序号	讲座名称	日期	课件下载
1	异构数字方志集成平台——数字方志知识集成与分析	2019-3-6 13:00-14:30	下载
2	Tableau——让数据可视化如此简单	2019-3-13 13:00-14:30	下载
3	ANTNeuro脑电系统——脑电信号采集与分析	2019-3-20 13:00-14:30	下载
4	眼镜式眼动追踪设备——眼动行为数据捕捉与分析	2019-3-27 13:00-14:30	下载
5	桌面式眼动追踪设备——眼动行为数据捕捉与分析	2019-4-3 13:00-14:30	下载
6	ArcGIS——空间数据的完美呈现	2019-4-10 13:00-14:30	下载
7	人文社会科学大数据平台——人文社科数据汇聚加工厂	2019-4-16 13:00-14:30	下载
8	incoPat——全球专利数据库平台	2019-4-24 13:00-14:30	下载
9	全球视野下的数字人文：研究与实践（地点待定）	2019-4月思想节日	下载
10	学术地图发布平台的应用	2019-5-15 13:00-14:00	下载
11	人文社科数据联盟链平台——联盟数据活动的“记录者”	2019-5-22 13:00-14:30	下载
12	星环极速大数据平台——社科数据共享存储与挖掘分析	2019-5-29 13:00-14:30	下载
13	DHAIP——数字人文人工智能平台	2019-6-5 13:00-14:30	下载
14	海外中国学研究者专题数据库	2019-6-12 13:00-14:30	下载
15	民国学人专题数据库	2019-6-19 13:00-14:30	下载
16	上海高新技术企业专题数据库	2019-6-26 13:00-14:30	下载

图9 华东师范大学闵行校区图书馆系列数字人文讲座

3.3.2 数字人文沙龙

面向特定群体开展工作坊：联合中文系师生，有针对性的举办工作坊，介绍数字人文工具的同时，征求中文系师生的数字人文需求，力求开展数字人文服务协同合作，使图书馆员参与到具体的数字人文研究活动中，成为数字人文成果的创建者。

在以后的工作中，华东师大正在考虑从三方面入手：一，与院系教师合作，聘请专业教师开展数字人文工具培训。二，加强学科馆员数字人文技能培训，使图书馆可以更好地翻译、整合数字技术和人文学者之间的对话，服务人文学者。三、尝试针对不同的读者开展初级科研工作坊和高级科研工作坊，初级工作坊，旨在培养学生从事科研活动的基础知识和工具使用。高级工作坊旨在培训学生从事数字人文深层研究活动的工具使用，帮助高年级学生进行具原创性学术研究。

4 结语

图书馆为各类科学教学科研提供专门的、专业性的优质服务是其重要任务之一。数字人文研究的活跃，对图书馆的人文服务提出了新的要求。高校图书馆数字人文服务建设和数字

人文中心建设的根本区别是在于，高校图书馆建设是在现有图书馆资源建设的基础上，通过充分挖掘现有图书馆内部资源进行整合优化，为本校人文教学研究服务。所以高校图书馆数字人文服务的建设需要处理特殊服务与一般服务的矛盾，即不能以偏盖全，偏离图书馆的主要职能，也不能脱离数字人文的客观性需求，无所作为。高校图书馆承担着文献收集整理、开展咨询服务以支持本校教学研究需要的根本职能，所以我们在数字人文服务模式的研究主要围绕图书馆在资源建设、信息素养咨询培训等隶属于图书馆本职服务等方面展开的。取得如下成果：

数字人文服务对数据资源的要求，已经需要图书馆从资源的被动提供者向信息资源的加工者转变。主要体现在对数据资源需要的信息粒度更细、数据分类更细、需要多数据融合支持等方面。所以国内图书馆需要在馆藏方面考虑增加更加专业型的数据库，还需要具备多数据融合的技术能力。树立图书馆在数据资源使用上的“只要我有，你必可用”形象。

信息素养教育咨询服务呈现主动服务和被动服务相结合，多渠道开展的模式，其核心要素就是要降低人文学者获取咨询服务的门槛。对图书馆而言，一是要建设专业的、有针对性的数字人文信息访问渠道，如数字人文专栏；二是要开展有针对性的、高质量的培训、讨论会等活动吸引人文学者走进来。

数据资源的应用呈现合作模式。图书馆在各类研究中，不只是初级数据提供者，而是可以发挥自身学科优势，成为科研项目的合作者，在科研项目中发挥更大效益。这些要求图书馆的需要提升自身的定位，提供鼓励科研的管理制度，促使馆员参加更多的项目中去。

参考文献

- [1] Ying Zhang, Shu Liu, Emilee Mathews, Convergence of digital humanities and digital libraries[M]. Emerald Group Publishing Limited. 2015
- [2] Karen Miller, Erik Champion, Lise Summers, et, Digital Humanities, Libraries, and Partnerships[M]. Elsevier Ltd. 2018.
- [3] CenterNet [EB/OL].[2018-1-18].<https://dhcenternet.org/>.
- [4] 于亚秀, 李欣. 美国高校图书馆的数字人文服务实践——以加州大学洛杉矶分校为例[J]. 图书馆论坛(9):30-37.
- [5] 首届北京大学“数字人文论坛”：跨界与融合：全球视野下的数字人文[EB/OL].[2018-1-13]. <http://www.lib.pku.edu.cn/portal/cn/news/0000001259>.
- [6] 第二届北京大学“数字人文论坛”：互动与共生-数字人文与史学研究[EB/OL].[2018-1-13]. <http://www.lib.pku.edu.cn/portal/cn/news/0000001486>.
- [7] 第三届北京大学“数字人文论坛”：孵化与实践-需求驱动下的数字人文项目[EB/OL].[2018-1-13].<http://www.lib.pku.edu.cn/portal/cn/news/0000001622>.
- [8] 潘雪, 陈雅. 国外高校数字人文课程设置结构分析——以 iSchools 联盟为例[J]. 数字图书馆论坛, 2017(10):68-72.