



基于 Omeka 与 IIIF 的特藏资源库建设研究与实践

——以华东师范大学近代教科书数据库为例

□ 张毅 陈丹*

摘要 数字阅读的流行以及商业搜索引擎与移动 APP 的快速增长,不仅挑战着图书馆作为知识文化中心的地位,也减少了图书馆高价值收藏的关注度,让读者迷失在鱼龙混杂的信息当中。此文在分析了国内“985 工程”高校图书馆特藏资源库建设的基础上,以华东师范大学图书馆近代教科书数据库为例,借助国际图像互操作框架 IIIF、开源特藏资源管理系统 Omeka 以及腾讯 AI 接口,重构图书馆特藏资源库建设思路,寻找图书馆无缝融合全球海量特藏资源的能力。

关键词 特藏 近代教科书 Omeka IIIF 人工智能

分类号 G250.76

DOI 10.16603/j.issn1002—1027.2021.03.014

身处信息爆炸时代,图书馆作为知识中心的地位不断受到来自搜索引擎、商业数据库以及社交网络的挑战。虽然图书馆保存有大量高价值的特藏资源,但传统的纸质资源服务模式已经无法满足读者获取信息的需求。缺少有效的数字资源管理、组织与揭示的系统与标准,只会导致数字特藏资源埋没在信息孤岛当中^[1]。国内高校图书馆自建数据库常用的商业系统软件有 TPI、麦达、TRS 以及 DIPS 等^[2],这些价格昂贵的商业软件使得经费紧张的图书馆无法承担建设费用,而且其加密的软件源代码与封闭的系统导致资源无法重用。由于软件开发成本的制约和资源揭示手段的快速发展,商业数据库软件更新速度远落后于信息技术的发展,在资源开放获取、知识图谱、跨平台访问以及自然语言处理方面,都无法满足读者的需求。虽然图书馆更加了解读者需求,但缺少技术储备的单个图书馆很难开发出具有号召力的特藏资源管理系统。

开源内容管理软件与国际合作社区的出现,给图书馆实现特藏资源管理、发布以及揭示提供了全新可能。本研究将选取具有极高社区活跃度的国际图像互操作框架 (International Image Interoperability Framework, IIIF) 来实现特藏资源库的建设^[3],并结合对 IIIF 支持较全面的 Omeka 资源管理软件作

为内容管理平台^[4],构建全新的特藏资源管理、发布与开放共享平台。随着 IIIF 标准被各国特藏资源管理机构所接受和认可,创建于 2006 年的 Omeka 也得到了广泛的关注。从 2019 年开始,辽宁大学图书馆的宫平、福建师范大学的邱建华以及厦门大学的陈晓亮都关注到 Omeka 系统在特藏资源与数字人文方面的重要价值^[5—7]。

以社区驱动的 IIIF 可以自由申请加入^[8],IIIF 社区成员创建了活跃度极高的谷歌讨论组^[9],可以在聊天群组 (IIIF Slack) 上分享自己的观点^[10],也可以在 GitHub 上获取社区提供的各种插件与案例^[11],成员可以在 IIIF Community Calendar 中跟进每周的电话会议计划^[12],还可以在 IIIF Google Drive Directory 中查看相关历史文件^[13]。目前国内也已经建立了微信群“IIIF-China”以及视频分享平台哔哩哔哩的 IIIF-China 账号^[14],鼓励国内 IIIF 社区成员积极交流。Omeka 与 IIIF 提供了非常友好的人工智能调用组件,便于特藏数据库调用人工智能接口。

1 高校图书馆特藏数据库调查分析

1.1 调查

为了全面把握我国特藏资源库建设现状,本研

* 通讯作者:陈丹,ORCID:0000-0002-7384-3530,邮箱:dchen@library.ecnu.edu.cn。



究于 2020 年 12 月 8—16 日, 调研了“985 工程”高校图书馆主页中的特藏资源栏目, 在 39 所“985 工程”高校图书馆网站中查找“特藏资源”“自建数据库”以及“特色资源”等类别, 并以检索式“insite: lib.*.edu.cn 特藏 or 自建 or 特色”在百度以及必应搜索引擎中检索以避免遗漏, 其中“*”表示对应高校。调研结果发现特藏数据库的建设与推广成为共识, 主页中有特藏资源栏目的图书馆有 27 所, 占 39 所高校图书馆的 69%, 数量共有 191 个特藏数据库, 可以看出, 重视特藏资源建设已经成为高校图书馆的共识。特藏数据库数量较多的图书馆有北京师范大学图书馆、华南理工大学图书馆、南开大学图书馆以及北京大学图书馆。

1.2 存在的问题

虽然高校图书馆已经意识到特藏资源的重要价值, 但是在特藏资源管理、揭示以及开放共享方面还有较大欠缺, 表现在以下几个方面:

(1) 大部分特藏资源仅服务于本校读者

在 27 所图书馆中, 2/3 的图书馆特藏资源是部分面向校外读者开放浏览权限, 1/3 则是所有特藏资源库都必须通过 VPN 才可以访问, 这也是当前国内特藏资源服务的困境, 资源一旦只面向本校读者服务, 就很难扩大影响力, 也难以进一步倒逼数据库服务质量的提升, 但资源开放后又可能会出现违反版权与使用规定等方面的问题。

(2) 缺乏开放共享标准与开源软件

在调研的 27 所图书馆中, 提供的资源开放共享方式有 3 种, 分别是 CALIS 规范、IIIF 以及元数据下载。在这些提供资源开放共享的数据库中, 只有厦门大学学术典藏库是基于开源软件 DSpace 构建^[15]; 利用 IIIF 开放接口的只有复旦大学图书馆开发的印谱数据库^[16], 其 IIIF 技术部分由上海慧游文化传播有限公司提供; 其中, 符合 CALIS 规范的特藏资源库的图书馆有 8 所, 可以实现特藏资源在元数据层面的开放共享, 国内主流的自建数据库软件基本都已经遵循 CALIS 规范, 符合 CALIS 规范的数据库在经过复杂的数据校验后, 需要在专业人员干预下, 实现元数据的批量收割, 资源也只能单方面地提交给 CALIS 中心, 这种开放共享模式不够灵活, 无法实现资源的细颗粒度开放共享。

(3) 展示形式单一

特藏资源主要以图片与 PDF 两种方式存在, 受

限于外网访问限制, 仅从开放外网访问的资源中分析发现, 大多数数据库仅支持简单浏览, 只有个别特藏资源支持虚拟展览、缩放、旋转以及灰度等访问方式。

(4) 鲜有数字人文工具的应用

知识图谱、可视化以及时空展示已经在图书情报领域有较多研究与实践, 但是从调查的结果看, 图书馆在特藏资源揭示方面很少用到这些数字人文方法。

(5) 不适合移动设备浏览

从调查结果看, 可以在移动设备上访问的特藏数据库屈指可数, 在当前这个移动互联网时代, 不能适配移动设备, 会严重影响读者随时随地获取资源, 无法充分发挥特藏资源的价值。

(6) 没有采用与时俱进的人工智能技术

当人工智能技术已经像水和电一样成为一种可以轻松获取的基础设施时, 特藏资源库应尽快采用人工智能技术, 对包含的图片、语音进行识别分类, 便于读者发现。

2 华东师范大学近代教科书数据库需求分析

对“985 工程”高校图书馆特藏数据库建设现状的调查分析发现, 虽然图书馆界已经意识到特藏资源数据库是信息爆炸时代体现图书馆独特价值的重要载体, 也已经建成了大量的特藏库对读者开放, 但是仍旧存在信息孤岛、受商业数据库软件限制、不适合移动设备以及资源揭示不够深入等问题。在吸取了国内特藏资源库建设经验的基础上, 本研究将借助开源软件与国际合作社区的力量, 以华东师范大学近代教科书数据库建设为例, 探索人工智能时代, 适合移动互联网环境, 具有开放共享能力的特藏资源数据库建设方案。

2.1 近代教科书资源现状

由于近代图书纸张材料的制作工艺, 即使采用非常苛刻的保存环境, 其破损速度也非常快, 华东师范大学图书馆已经不再提供近代图书的借阅服务。为了挽救近代图书, 发挥其价值, 图书馆优先考虑数字化近代图书, 并购买专业存储设备保存。在已经数字化的近代特藏资源当中, 近代教科书约有 2800 多种, 收藏的出版年代最早的教科书是清朝光绪二十二年(1896 年)的中学几何教科书《中等教育克依其氏最新平三角法教科书》, 收藏教科书最多的年份



是 1933 年前后, 详见图 1, 出版地主要集中在上海和北京, 出版地在上海的就超过了其他城市之总和。主要的出版社是商务印书馆、中华书局以及世界书局等。除了师范用书收藏量较少外, 其他学龄段教

科书数量差距不大。而学科的差距比较大, 从统计的结果看, 近代教科书学科覆盖非常全面, 其中收藏数量较多的有算数、国文、自然、中国历史等学科, 详见图 2。

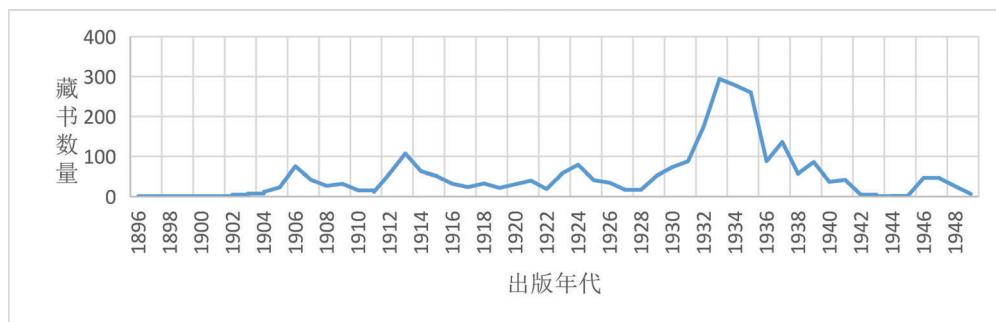


图 1 华东师范大学图书馆收藏近代教科书的出版年代分布



图 2 华东师范大学图书馆收藏近代教科书的出版信息与学科分布

2.2 数据库功能设计

根据读者在移动互联网与人工智能环境下获取资源的习惯, 近代教科书数据库提供泛在化跨平台的服务功能尤其重要, 人工智能技术的应用则可以根据读者的使用习惯, 个性化推荐资源。强大的资源揭示与管理能力, 是近代教科书资源发挥自身价值与长期保存的必要条件; 系统的开放性则可以为近代教科书数据库提供强大的生命力, 使其融入全球开放资源统一发现系统。详细的数据库功能如图 3 所示。

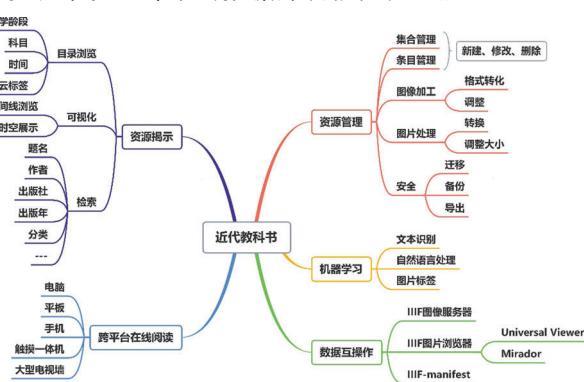


图 3 华东师范大学图书馆近代教科书数据库功能

(1) 跨平台在线阅读

页面自适应与图片动态加载技术的应用, 为近代教科书数据库使用多种在线阅读设备提供了技术保障, 使得数据库可以根据显示器的大小自动匹配合适的页面, 并为读者提供资源的深度缩放阅读服务。

(2) 机器学习

机器学习不仅可以轻松对图片中的文字进行识别, 而且可以对图片进行分析、标注、分类以及个性化推荐, 运用到的技术主要有机器视觉与自然语言处理等。

(3) 资源管理

近代教科书包含着海量图片资源, 需要有一款功能强大的资源管理系统, 对近代教科书中的元数据与图片进行精细化管理、备份、动态格式转化以及迁移, 并匹配国际通用的元数据格式。

(4) 资源揭示

对近代教科书元数据进行细颗粒度分析, 从时间、空间、分类等多个角度揭示特藏资源。在系统功能上, 为读者提供资源高级检索、目录浏览以及可视化工具, 来提升系统的资源揭示能力。



(5) 数据互操作

国际图像互操作框架的标准及其社区, 为实现近代教科书资源的互操作提供支持, 使得资源可以在符合 IIIF 接口的系统之间自由整合。

2.3 数据库系统架构

近代教科书数据库系统整体架构可分为 4 层, 包含 5 个模块, 如图 4 所示。最底层是数据层, 以华东师范大学图书馆已经数字化的近代教科书为基础, 同时包含了与其他师范类高校图书馆交换获得的电子版教科书。由于采用了 IIIF 接口, 任何以 IIIF 发布的近代教科书都可以无缝嵌入到本系统中, 共同构成了近代教科书的数据基础; 第二层是 Omeka 特藏资源管理系统, 这一层负责近代教科书的增加、删除、发布等功能, 同时还需具备图像转化、编辑等功能, 整个近代教科书数据库的检索、发现也是由这一层实现; 第三层是在 Omeka 的基础上实现 IIIF 的图像接口与展示接口, 这一层是近代教科书数据库系统的核心, 机器学习可以嵌入在 Omeka 系统中, 识别每一张图片的内容, 并以注释的形式在 IIIF 展示接口中显示; 最后一层是发布层, 发布需要做到适配电脑、手机、平台等不同的设备, 还需要利用时间轴、云标签等手段更好地展示资源, 以及提供分享链接等功能。



图 4 华东师范大学图书馆近代教科书数据库系统架构

3 近代教科书数据库系统实施

对华东师范大学近代教科书资源原始数据进行梳理是实施的第一步, 之后需要部署资源管理与发布的 Omeka 系统, 分析 Omeka 系统存储文件的目录格式以及数据库中的数据, 批量导入本地教科书到 Omeka 系统; IIIF 接口图像服务器的配置与展示接口的开发, 是近代教科书资源实现开放共享的基础; 实施的最后一步则是调用腾讯云人工智能接口实现图片文字识别与分类等功能。

3.1 Omeka 部署

(1) 系统环境部署

Omeka 系统包含 Omeka S、Omeka Class 与 Omeka.net 三种版本, Omeka S 功能更多, 但部署与运维较复杂, Omeka.net 是在线版本。综合考虑之后, 最终选择 Omeka Class 作为近代教科书数据库系统的管理平台。Omeka Class 推荐运行在 linux 服务器上, 近代教科书数据库采用 Ubuntu 14 作为其操作系统, 部署 Apache 作为其 HTTP 服务器, Mysql5.0 作为数据库服务器, PHP5.4 作为脚本语言运行环境, 对图片的加工采用 ImageMagick 工具。硬件服务器通过华东师范大学图书馆的虚拟化环境提供, 配置为 1TB 硬盘, 32GB 内存, 4 路双核 CPU 处理器。

(2) Omeka 安装

将下载的最新版本 Omeka 系统上传到 Web 服务器的根目录, 在浏览器中输入 Web 服务器的地址 (<http://xxxxxx.xx>), 启动安装程序, 在此之前需要先修改 Omeka 的“db.ini”文件, 修改其中的数据库地址与端口, 然后根据系统提示操作, 完成 Omeka 的部署安装。在服务器路径后面加 admin (<http://xxxxxx.xx/admin>), 进入后台管理界面, 配置 Omeka 系统网站的名称、简介、安全性等内容, 完成近代数据库系统的配置。通过创建集合、条目上传文件, 就可以通过 Omeka 系统管理与发布近代教科书资源。

3.2 IIIF 功能实现

IIIF 最重要的两个功能是实现图片处理的图片服务器与为用户提供良好用户体验的浏览器, 这两个方面的功能都可以通过 Omeka 官方提供的 Universal Viewer 插件实现, Universal Viewer 是专门针对图片电子书在线阅读的 IIIF 插件, 由丹尼尔·贝尔特罗 (Daniel Berthureau) 开发设计^[17], 内置 IIIF 图片服务器^[18], 不仅可以对图片进行转化修改, 其动态图片加载技术, 还可以为用户提供超高清图片的在线浏览器与深度缩放服务。

在 Omeka 系统中安装 IIIF 插件非常容易, 通过 Omeka 官网下载 Universal Viewer 插件, 上传到本地 Omeka 安装目录中的 Plugins 下, 登录到本地 Omeka 系统后台, 打开插件管理界面进行安装配置, 需要设置 Universal Viewer 浏览器作为图像服务器, 选择需要显示 Universal Viewer 浏览器的页面位置。



3.3 近代教科书资源批量导入 Omeka 系统

(1) 图片上传与处理

已经数字化的 2800 本近代教科书, 图片总量约 35 万张。手动上传如此海量的资源显然无法实现, 因此本研究通过分析 Omeka 系统存放文件的格式、路径以及数据库的配置, 最终采用批量上传方式, 将 2800 本电子书一次性上传到服务器。Omeka 系统的所有资源都存放在其“files”文件夹下面的 4 个目录中, 分别是 fullsize、original、square_thumbnails、thumbnails。在通过 Omeka 后台上传图片时, ImageMagick 工具会对上传的图片进行加工, 全尺寸的 JPG 格式图片存放在 fullsize 目录下, 方形缩略图和缩略图分别存放在 square_thumbnails、thumbnails 下面, 而源文件则存放在 original 目录下。当读者访问 Omeka 数据库时, 系统会从这 4 个路径中读取图片, 并显示给读者查看。根据 Omeka 系统的文件存放方式, 将所有近代教科书资源的源文件上传到 original 目录, 然后利用命令行调用 ImageMagick 对 original 目录的图片进行批处理, 生成 fullsize 需要的图片格式与尺寸, 用到的 ubuntu 命令行是: `ind-type f-exec bash-c '[[-n" $(file -bi "$1" | grep tif)"]] && convert-resize 40%-quality 15% "$1" jpg: "$1.jpg"-{}\'`; 缩略图和方形缩略图的生成方式相同, 这样就在 Omeka 系统中准备好了近代教科书的所有文件资源。

(2) 元数据导入

导入元数据与每张图片的物理地址到 Omeka 的 Mysql 数据库中, 并做好映射关系后, 即可完成数据的批量导入。Omeka 数据库中与批量导入有关的表有 7 个, 分别是 omeka_collections、omeka_items、omeka_element_texts、omeka_files、omeka_records_tags、omeka_tags、omeka_search_texts。其中, 近代教科书数据库的集合信息存放在 omeka_collections 中, omeka_items 存放近代教科书的条目信息, omeka_element_texts 存放每条目录及集合的中文名称、作者、出版年等元数据信息, omeka_files 存放每本教科书包含的图片物理地址, omeka_records_tags、omeka_tags 存放标签名称与每个标签对应的具体图书, 可以检索到的字段都存放在 omeka_search_texts 当中, 为了提供更加详细的检索信息。在这些表中, item_id、collection_id 对应表

的主键, 作用是将多个表的内容关联起来。

3.4 整合国外近代教科书资源

国际上已经有大量的特藏资源以开放共享的 IIIF 标准发布, 其中不乏非常稀有的近代教科书资源, 查找这些资源并添加到系统, 可以为读者提供更加丰富的资源。Omeka 调用 IIIF 标准发布的特藏资源也非常容易, 只要将 IIIF 的 manifest 网址添加到近代教科书的图片浏览器插件中, 读者就可以像浏览本地资源一样来访问全球的近代教科书资源。Omeka 系统还专门提供了将 IIIF 资源导入本地的插件 IIIF toolkit, 使得导入外部近代教科书资源到系统成为一件非常容易的事情。

国际上基于 IIIF 标准建设的特藏资源数据库都是以开放共享为宗旨, 目前已经有一些 IIIF 资源集成发现平台, 为查找国外近代教科书资源提供有力工具, 基于 IIIF 标准的资源又可以无缝嵌入到华东师范大学近代教科书数据库, 丰富近代教科书数据库的内容。常见的 IIIF 发现系统有 iiif-universe、iNQUIRE、Biblissima IIIF-Collections、Europeana's IIIF filtered search 以及 OCLC's IIIF Explorer 等^[19-23]。在这些发现系统中以“textbook”为检索关键词, 时间范围选择 1949 年之前, 就可以检索出大量高质量的稀有教科书资源。

3.5 人工智能接口调用

人工智能接口的调用有两种方案, 一种是使用 Omeka 自带的 Text Analysis 文本分析插件^[24], 即利用 IBM 公司的自然语言处理接口 (Watson Natural Language Understanding, NLU) 进行文本分析^[25], 由于对中文进行文本分析的功能不完善, 而且不能进行图像文字识别与图像标签, 所以最终选择第二种方案, 即国内的腾讯人工智能接口^[26]。腾讯人工智能接口的 API 密钥包含 SecretId 与 SecretKey 两个字段, 他们是驱动腾讯人工智能接口的关键, 根据近代教科书数据库系统的软件开发语言, 选择安装最新版本的 PHP 腾讯云 SDK 3.0。接口的调用比较简单, 近代教科书数据库用到的接口分别是通用印刷体识别 (AdvertiseOCR)、图像标签 (DetectLabel) 以及关键字提取 (KeywordsExtraction) 接口。以近代教科书《新中华音乐课本教授书》中的一张图片为例, 调用图像标签接口的 PHP 代码与运行效果如图 5 所示, 从识别结果可以看出, 机器发现了图片中包含人物。

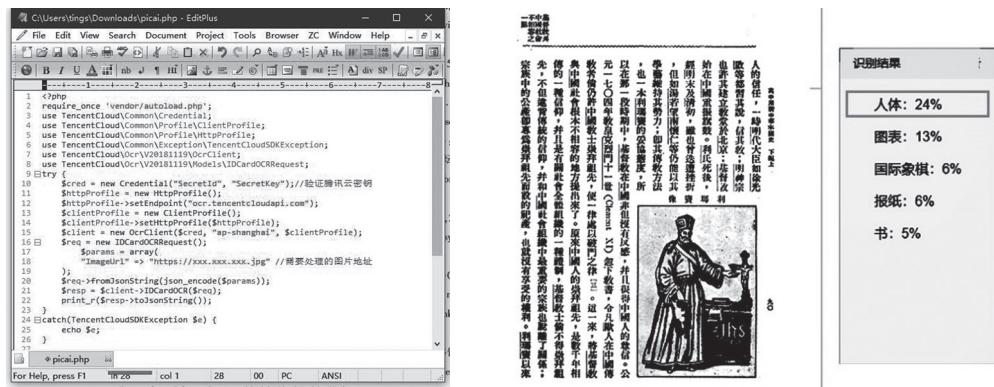


图 5 华东师范大学图书馆近代教科书数据库调用腾讯图片识别接口代码与运行效果

3.6 近代教科书数据库用户界面

Omeka 系统自带多种外观主题,可以从 Omeka 官网下载使用,也可以自己开发。经过项目组的测试,选择了最适合电子书显示的“Thanks, Roy”主题^[27],它具有界面简洁大方与响应速度快的特点,用户可以修改主题页面的 LOGO、网页底部文字、页面颜色、条目以及集合的显示等,非常灵活。页面导航方面,项目组根据近代教科书内容特点,以小学、初中、高中三个学段进行大的分类,每个学段又根据学科进行分类,方便读者查找感兴趣的内容。

云标签是读者迅速查找需要资源的有效手段,近代教科书数据库利用人工智能的图像标签接口、文本关键字提取接口获取每张图片的分类数据,并结合每本电子书的元数据内容,进行标签云的生成,读者可以通过标签云深度定位到自己感兴趣的内容。简单搜索与高级搜索相结合,方便读者通过关键字定位到一本书的每一页,简单搜索以关键字模糊匹配来发现读者需要的资料,高级搜索可以选择元数据中的特定字段进行交叉检索,检索方式灵活多样。



图 6 华东师范大学图书馆近代教科书数据库的 4 种浏览模式效果

3.7 服务方式

基于响应式页面设计的近代教科书数据库可以在电脑、手机、平板电脑上都获得统一的显示效果。利用 Universal Viewer 电子书浏览器的图片动态加载与深度缩放特征,在移动终端上,读者也能获得较好的阅读体验。Universal Viewer 还提供全屏浏览模式,以类似电脑客户端或手机应用的形式为用户提供近乎完美的深度阅读体验。借助全屏模式,还可以将所有的电子书都加载到 Universal Viewer 浏览器中,使得近代教科书数据库在大型触摸一体机上也具有良好的呈现效果。不同浏览模式的效果如图 6 所示。

近代教科书数据库以 IIIF 方式发布资源,通过网页上的 IIIF-manifest 连接,可以直接将资源无缝嵌入到其他 IIIF 系统中。对于没有 IIIF 的系统,还可以通过 Web Widget 方式,将近代教科书数据库嵌入自身系统中。除了以 IIIF 方式开放资源,Omeka 系统自身也提供了 Omeka-xml、dcmes-xml 等方式开放数据,供其他系统调用。



4 总结与展望

结合开源内容管理系统 Omeka、国际图像互操作框架 IIIF 以及机器学习技术构建的近代教科书数据库系统,给华东师范大学图书馆在特藏资源管理与揭示方面打开了一条全新的思路,摆脱了特藏资源系统长期依赖软件供应商的被动局面,参与到国际特藏资源保存与共享当中,为图书馆在其他特藏资源管理、发布、揭示与共享方面积累了理论与实践经验。基于响应式页面设计的近代教科书数据库可以完美自适应多种终端,可全屏显示的电子书阅读器配合图片动态加载技术,成为读者碎片化学习的有力工具。华东师范大学图书馆近代教科书数据库是 IIIF 技术在国内图书馆特藏资源揭示方面的一次重要实践,希望能够为国内图书馆、档案馆、博物馆等机构在特藏资源数据库建设方面提供有益参考。

参考文献

- 1 王俊杰.冲出信息孤岛,实现数字资源共享[J].大学图书馆学报,2004(3):16—18.
- 2 黄晓斌,钱国富.数字图书馆开发平台及其发展趋势[J].中国图书馆学报,2005(4):53—57.
- 3 IIIF.About IIIF[EB/OL].[2019—10—10].<https://iiif.io/about/>.
- 4 Digital Scholarship. Omeka Classic[EB/OL].[2019—05—27].<https://omeka.org/>.
- 5 宫平.OMEKA 在图书馆数字人文项目中的应用[J].图书馆研究与工作,2019(10):92—95.
- 6 邱建华.Omeka 系统在数字人文研究中的应用剖析[J].情报探索,2019(10):104—109.
- 7 陈晓亮,苏海潮,刘心舜.图书馆特藏数据结构化的探索[J].图书馆杂志,2019,38(6):44—48,91.
- 8 IIIF.IIIF community[EB/OL].[2020—01—07].<https://iiif.io/community/>.
- 9 IIIF.IIIF discuss[EB/OL].[2020—01—10].<https://groups.google.com/forum/#!forum/iiif-discuss>.
- 10 IIIF.IIIF-slack[EB/OL].[2020—01—20].<http://bit.ly/iiif-slack>.
- 11 IIIF.IIIF community GitHub page[EB/OL].[2020—02—09].<http://github.com/iiif>.
- 12 IIIF.IIIF Community calendar[EB/OL].[2020—10—12].<https://iiif.io/community/groups/>.
- 13 IIIF.IIIF Google drive directory[EB/OL].[2020—08—18].<https://drive.google.com/drive/folders/0B9EeoRu2zWeraXpHNXpnZThUZVE?usp=sharing>.
- 14 Bilibili IIIF-China[EB/OL].[2020—06—15].https://space.bilibili.com/590634547?spm_id_from=333.788.b_765f7570696e666f1.
- 15 厦门大学图书馆.厦门大学学术典藏库[EB/OL].[2020—09—09].<https://dspace.xmu.edu.cn/handle/2288/72>.
- 16 复旦大学图书馆.复旦大学印谱数据库[EB/OL].[2020—10—27].<http://yin.fudan.edu.cn/>.
- 17 Berthureau D. Universal Viewer[EB/OL].[2019—10—19].<https://omeka.org/classic/plugins/UniversalViewer/>.
- 18 Pillay R. IIPImage[EB/OL].[2019—11—12].<https://iipimage.sourceforge.io/>.
- 19 The World Wide Graph.iiif-universe[EB/OL].[2020—10—01].<https://graph.global/universes/iiif/>.
- 20 The Bodleian Libraries. Digital Bodleian[EB/OL].[2020—01—19].<https://digital.bodleian.ox.ac.uk/>.
- 21 Biblissima.IIIF-Collections[EB/OL].[2020—10—09].<https://iiif.biblissima.fr/collections/>.
- 22 Europeana. Europeana's IIIF filtered search[EB/OL].[2020—03—07].https://www.europeana.eu/portal/en/search?q=provider_aggregation_edm_isShownBy%3A*iiif*&view=grid/.
- 23 OCLC.IIIF Explorer[EB/OL].[2020—05—12].<https://researchworks.oclc.org/iiif-explorer/>.
- 24 Roy Rosenzweig Center for History and New Media.Text Analysis[EB/OL].[2020—09—11].<https://omeka.org/classic/plugins/TextAnalysis/>.
- 25 IBM.Natural language understanding[EB/OL].[2020—10—13].<https://cloud.ibm.com/catalog/services/natural-language-understanding/>.
- 26 腾讯.AI 开放平台[EB/OL].[2020—07—17].<https://ai.qq.com/>.
- 27 Roy Rosenzweig Center for History and New Media.Thanks, Roy[EB/OL].[2019—03—16].<https://omeka.org/classic/themes/default/>.

作者单位:华东师范大学图书馆,上海,200241

收稿日期:2020 年 11 月 19 日

修回日期:2021 年 3 月 2 日

(责任编辑:关志英)

Research and Practice of Special Collection Database Construction Based on Omeka and IIIF

—Taking the Database of Modern Textbooks of East China Normal University as an Example

Zhang Yi Chen Dan

Abstract: The popularity of digital reading and the brutal growth of commercial search engines and mobile APPs have not only challenged the status of the library as the knowledge and cultural center, buried the high-value special collections of the library, but also caused readers to get lost in the mixed information. Based on the analysis of the construction of the special collection of the “985 Project” academic libraries and taking the modern textbook database of the East China Normal University Library as an example, this study adopts the international image interoperability framework IIIF, the open source special collection resource management system Omeka and Tencent’s AI interface to reconstruct the idea of constructing library special collection resources in order to seek the ability of libraries to integrate seamlessly global mass collection resources.

Keywords: Special Collections; Modern Textbook; Omeka; IIIF; Artificial Intelligence