

数字人文在 边疆历史研究中的应用

□ 陈建红 史话

东北师范大学 历史文化学院 吉林 长春 130024

一、围绕边疆研究建设数字人文平台

史料是历史研究的基础,系统的、充分的、经过检验的史料是历史研究的坚实基础。相比国内偏向传统的边疆史料整理,国外有关边疆历史的数字人文研究已经先行一步。在古籍数字化、文献数据通用共享、国际话语传播、边疆通识教育等方面已经取得了长足的发展,逐渐形成“传统学术研究+互联网技术体系+融媒体传播”的新技术条件下的学术研究与传播体系。

1. 与边疆有关的历史文献数字化发展迅速。在数字化方面,海外开始的较早,进展也比较迅速。保存中国边疆历史文献较多且数字化做得比较好的是美国、日本、韩国等国家。

2. 数字人文技术应用于边疆历史问题研究。海外数字人文技术在历史研究领域的应用并不局限于古籍、文献、档案资料的数字化,已经开始对学术资源进行数据处理和可视化呈现,并通过通用数据实现跨平台共享。例如,东北亚历史网,把中国、俄罗斯、朝鲜半岛的历史遗迹(考古成果)、古今地名(GIS数据)以及与之相对应的学术研究成果(文献、专著、论文)通过知识图谱技术建立数据关联,在其数据平台实现可视化呈现。韩国珍稀档案在通过瓦片地图实现高清浏览的同时,还将图片内容录入文本,在保持珍稀档案原图显示的同时实现全文多语种检索,并且通过标准数据(XML)实现数据通用。文本数据化打破了出版著作、档案、图片之间的“信息孤岛”状态,实现档案由数字图像信息到文本数据的升级,在保证信息传达的准确性的基础上,为研究者高效利用文献资源提供了可能。

很多海外机构都通过自己的门户数据网站,推动历史数据面向全网开放。美国国会图书馆几乎所有可见的人文数据都是开源的,日本、韩国大多数公共机构的人文资源也面向其国内用户开放。

文献通用化数据的开发与开放为我们获取和使

用这些数据提供便利的同时,也为中国边疆历史信息数字化提供了参考样例。当然,这也对研究者掌握数字人文技术,拓宽研究资源的获得渠道,改善历史研究与书写方法,进一步提升研究质量提出了新的数字人文技术要求。

3. 数字人文技术推动边疆历史话语体系国际化。数字人文支持下的边疆历史研究和话语体系传播呈现出国际化趋势。一是学术研究机构 and 文献平台在加强学术资源搜集、整理与编译,各自齐头并进保持特色的同时,更加注重彼此资源的互联互通,逐步形成立足自身,联合同行,面向全网的学术文献体系。这种体系既是资源共享的平台,更是历史话语传播的公共空间。二是为了克服语言障碍,便于历史话语传播,部分与边疆有关的数字人文项目采用双语、甚至多语种传播体系。三是海外积极利用数字人文技术优化历史数据表达。

由于传统的文本记述往往秉持详内而略外,与复杂的边疆历史所伴生的往往是文献记述缺如,有限的记述也表现为语言简约,史事简略,而疆内疆外的当代解读又往往彼此冲突,因而,相关的研究既需要扎实的理论建构支撑,也需要大量的史料实证,更需要科学数据的支持。数字人文技术在边疆历史研究领域的应用是历史研究发展的趋势之一。掌握数字人文技术,才能掌握更多边疆历史研究的资源,跟进研究趋势。无论是调取海外文献,研判学术动态,还是完善中国边疆历史话语体系建构,加强话语的海外影响,都需要引入数字人文技术。

二、新文科实验室在边疆历史 数字化建设中的作用

新文科实验室是深入推动多学科交叉,促进边疆历史数据资源开发和科研创新的有效路径。在具体的数字人文实践中,边疆史学者存在两种困境:一是历史学者缺乏技术背景,面对日益丰富的学术资源,如何使用大数据开展科研和教学停留在“学术设想”

层面,无力施行;二是计算机等技术专家缺乏边疆历史专业知识,前沿技术找不到具对性的专业方向,无的放矢。实验室充分利用学科优势,跨院系展开合作研究,实现文、理、工学科之间的优势互补和资源交叉。以实验室为载体发挥的资源整合优势,在教学特色、人才培养、科研创新方面获得较大的突破。新文科实验室在边疆数字化建设中具有以下几个方面的作用。

第一,新文科实验室有利于边疆数字人文平台的开发、升级与日常维护。数据平台的开发需要人文学者与技术人员深度沟通,通过反复实验不断修正架构、改进技术才能完成。例如边疆数字人文平台最早的计划是构建一套边疆学者可用的历史地理信息编辑与知识地图发布系统,以克服历史学者在历史地理考证中的专业局限,补充历史学者历史地图绘制技能的短板。在随后的科研实践中,发现研究者首先需要解决的是历史文献的查询效率以及文献引用准确性问题。于是,升级原有的历史地理信息架构,构建了一套以文本对照与编辑系统为核心的文献体系。在完成文献体系之后,又发现文献解释的多维性是文献准确性基础上的又一需求,平台增设了百科系统,以实现文献的实时增量解释。以实验室为载体的自主研发平台在日常运营维护和课题研究过程中就可以完成平台更新,使平台始终在解决历史研究前沿问题的过程中保持技术活力。

第二,新文科实验室可以通过构建技术软环境全程跟踪服务边疆历史研究。数据挖掘、文本计算、数据可视化等数字人文技术,需要通过编程实现,对于缺乏相应技术背景的历史研究者来说,存在一定的困难。大多数人文学科的研究选题都属于“个性化”问题,在技术需求上有一定差异。尽管部分技术模块可以通用,当研究深入到特定领域时,技术就需要做出相应的调整。新文科实验室可以全程辅助边疆研究。例如,我们在开展“《韩国文集丛刊》中的中国边疆史料辑录”项目时,首先要通过对开源数据处理获得韩国文集丛刊的全套数据;第二步,对文献题录进行数据编码;第三步,通过文本计算获取与中国边疆历史相关的内容,形成初步挖掘结果;第四步,通过数据清洗获取核心文本段落;第五步,通过专家系统验证,确保辑录信息的准确性和完整性;最后,利用历史学知识对获取数据进行分类。在这一过程中,数字人文工作团队需要不断地调整技术参数,以确保辑录结果的科学性和准确性。辑录结果校验完成后,还需要对所获取的文本进行分析,以可视化的方式呈现计算结果,以方便研究利用。研究过程中,数字与人文紧密结合,构建文本计算模型,而模型迁移到其他类似项目中,也需要根据项目特点和研究需求进行必要的调整,这些都需要新文科实验室的辅助。

第三,自建新文科实验室更有利于边疆数据积累和数据安全。历史研究过程的积累往往和研究结果一样具有价值。过程的积累是进一步深入研究的基础,一个好的研究过程往往是一个新研究方向的窗口。借助外在平台进行的数字人文研究,最大的弊端是不能完整地保存研究过程,即便是保留研究进程的相关数据,也不利于在特定研究节点开辟新的研究方向,自建实验室则可以避免出现类似问题。实验室的所有实验环节都有严谨记录,算法模型和数据都得到完整的保留,方便研究结果复检的同时,也方便研究者以特定节点为起始,开启新的研究选题。自建实验室也可以为数据安全提供保障。边疆历史研究的数据安全至关重要,研究项目数据泄露,会产生不可预测的风险。无论是使用基于互联网的通用数字人文平台,还是使用其他形式的数字实验室,数据安全都存在隐患。自建实验室有规范的数据备份制度和相对成熟的数据安全保障机制,可以实现敏感数据的封闭式管理,在相当程度上保证数据安全。

第四,团队式运作,有利于科研的良性生长与学术创新。文科实验室是文科实验、实践、实训的重要载体,是培养创新型文科人才的重要平台。实验室建设涵盖了人文社会科学领域内多个学科的交叉、融合、渗透或拓展,是人文社会科学与自然科学交叉融合形成的文理交叉、文工交叉等新兴领域,这种团队式运作给予科研以极大的延展空间。边疆问题的研究存在很多不确定性,许多新问题会随着国际形势的发展而发生改变,问题解决并无固定模式,需要在实践过程中不断探索调整、完善。实验室正是这种充满挑战与变数的科研过程的承载平台。基于实验室平台的深入交流,可以让人文学者快速融入数字人文领域。同样,基于边疆需求的技术开发,也能够促进信息技术发现新的研究方向,实现技术创新。

综上所述,随着数字人文技术的快速发展,资源数字化程度不断提高,数据开源和体系融合成为数字人文平台发展的基本趋势。随着边疆研究资源的增加,数据的获取路径更加多元,边疆问题的研究视野得到拓展,解决问题的方法也随之增加,边疆历史研究迎来了新的发展机遇。但数字人文技术本身又在一定程度上局限了人文学者对这种机遇的认知和掌控。人文学者主导下,以边疆研究实际需求为导向,在数字化研究实践中开发一个贴近操作实际的边疆数字人文平台,可以成为边疆研究数字化的助力。在这一平台中,新文科实验室是技术依托,全球边疆数据整合是资源基础,历史地理信息和文本数据可视化既是边疆边界问题研究的表达助力,也是边疆知识普世化传播的途径。

■ 《云南师范大学学报(哲学社会科学版)》
2021年第4期,约12000字