

# 东南亚华人民间历史文献与数字人文

■ 丁荷生

东南亚华人民间历史文献的数量庞大而丰富,目前学界掌握的相关文献仅是“冰山一角”。许多材料在各地寺庙、会馆、家族与私人手中散落保存,尚未被系统收集、整理或数字化,甚至面临损毁与消失的风险。大部分文献出自福建、广东移民的传统书写体系,为研究东南亚华人社会、文化与经济活动提供了极其珍贵的第一手资料。

中国各地拥有数量庞大的民间文献。例如,国际敦煌项目(International Dunhuang Project, IDP)已将大量文献数字化,并计划在光学字符识别(optical character recognition, OCR)技术进一步完善后,全面开放全文检索系统。被称为“第二个敦煌”的徽州文书,自宋元以降累积数量或逾百万件,但因长期散藏于数十家公藏机构及私人手中,其内部逻辑系统遭到严重破坏。未来一项重要的学术任务,便是借助数字化的方法重新整合徽州文书体系,使其恢复历史原貌。除此之外,福建各地亦陆续发现大量民间档案,如永泰文书与其他八闽文献等,均是反映地方行政、宗族关系、祭祀活动及土地权属的珍贵史料。

民间文献的种类繁多,文类内部的流传方式亦不尽相同。这些文献主要包括碑刻、族谱、契约、账簿、仪式科本、戏剧文本等,既有在宗族内部传播的文献,也有在村落、庙宇或行业组织之间流通的档案,其中更包括只在师徒体系中才得以传授的内部资料。这些文类同样出现在东南亚华人社团中,但在当地又形成了特殊的发展轨迹,例如会馆的会议记录、行业组织的规约、侨民之间往来的批信、寺庙的募款碑刻等。侨批文献的数量极为庞大,涉及跨区域金融、社会网络与家族维系,为研究跨境移民史提供了细致的日常记录。

随着东南亚社会逐步进入高度数字化时代,许多旧式账簿、手写簿册与寺庙档案逐渐被废弃或毁损,抢救性的收集与数字化工作刻不容缓。而在庞大的材料面前,数字人文方法已成为不可或缺的工具。文本数字化、文本编码倡议(text encoding initiative, TEI)、地理信息系统(geographic information system, GIS)、社会网络分析(social network

---

丁荷生:新加坡国立大学中文系

analysis, SNA)、OCR、多媒体数据库等,都为民间历史文献的系统化处理提供了新的可能。从学术数据库的发展亦可看出这一趋势:西方古典学领域已建立如EAGLE Portal、Pleiades、ORBIS、FASTI Online等大型开放平台;汉学界也有中国历史地理信息系统(China Historical Geographic Information System, CHGIS)、中国历代人物传记资料库(China Biographical Database, CBDB)等数据库。类似的数字平台也正在东南亚华人研究领域逐步建立,为区域史研究带来跨学科的工具整合。

## 一、海洋联系:从福建碑铭出发

早期海上移民的商贸活动与扎根东南亚的历史,从各类史料中均可发现线索。例如,1138年的《祥应庙记》便记述了一位泉州船长远航三佛齐并获丰厚利润的故事,为研究中世纪海上贸易提供了珍贵证据。进入明初,郑和下西洋期间,《明实录》多次记录华人社群在苏门答腊、越南、马六甲及爪哇的活动。这些信息与当地出土的华人碑刻、庙宇遗迹相互印证,使我们能够重建早期海上移民的活动脉络。至17世纪中叶,马六甲的青云亭、巴达维亚的金德院、越南会安的多座会馆等,都显示福建人、潮州人、客家人等群体已在南洋扎根。

福建地方社会的碑刻研究,是理解海洋联系与跨区域宗教网络的重要基础。过去二十年来,厦门大学的郑振满教授与笔者陆续整理、出版了兴化(今福建省莆田市)、泉州、漳州等地的数千方碑刻。其中相当部分直接提及东南亚华人的捐助,显示明清以降福建本地宗教活动与海外移民社会之间的互动密切。碑文不仅记录了捐献人名与金额,更呈现跨海宗教组织所形成的分香网络、回乡进香活动与跨区域信仰体系的运作方式。

福建沿海地区在清代海禁与迁界政策下受到严重冲击,大量寺庙遭毁。碑刻记载南洋华侨在重建祖庙的过程中发挥了关键作用,如1696年厦门龙池岩寺派僧人赴巴达维亚(今印尼雅加达)募款,1698年巴达维亚保生大帝庙捐出两年香油钱支援清礁祖庙的修复等。这些记录显示福建神明的信仰网络往往依靠东南亚商业社群的资金维系。宗教网络的结构在地方碑刻中表现得高度清晰,各类祖庙详细记录东南亚分庙的名单、进香年代、捐献人群与跨海互动。改革开放以来的宗教重建更依赖东南亚华侨的捐助,大量寺庙、祠堂、佛寺、道观在海外资金的支持下重建或修复,形成跨海宗教复兴的重要组成部分。

如此密集的跨区域互动不仅体现在宗教机构层面,也体现在庙宇、会馆与家族档案的保存实践中。我们在新加坡、马来西亚、印度尼西亚、缅甸等地进行的田野调查显示,当地仍保存着大量尚未整理的碑刻、会议记录、账簿、祠堂档案与庙宇文书。这些材料对于研究东南亚华人社团的内部组织、仪式实践、经济网络与跨海交流具有极为关键的价值,亟须系统性的记录、数字化与学术发布。

## 二、东南亚华人碑刻的整理、出版与数字化

学界对于东南亚华文碑刻的系统研究可追溯到20世纪中叶。早在1957年,陈荆和便在《新亚学报》发表研究成果,利用越南会安的华人碑刻与会馆文献,分析17世纪、18世纪唐人街的发展与商业活动。<sup>①</sup>这是最早将南洋碑刻作为系统史料加以使用的研究之一。其后,饶宗颐于1970年出版《星马华文碑刻系年》,<sup>②</sup>对马来亚和新加坡华人碑刻的年代与史料价值做出重要评述,进一步奠定了碑刻研究在海外华人研究中的地位。陈荆和、陈育崧于1970年出版的《新加坡华文碑铭集录》,<sup>③</sup>以及庄钦永于1998年出版的《马六甲、新加坡华文碑文辑录》,<sup>④</sup>共同构成东南亚华人碑刻研究的基础文献。

在这项研究传统中,最具规模与影响力的,当属傅吾康(Wolfgang Franke)及其团队在20世纪八九十年代进行的大规模调查。他们先后在马来西亚、印度尼西亚与泰国搜集寺庙、会馆、宗祠与墓地碑刻,最终形成三套具有里程碑意义的大型碑刻汇编:《马来西亚华文铭刻萃编》《印度尼西亚华文铭刻汇编》与《泰国华文铭刻汇编》。<sup>⑤</sup>此外,黄文斌关于马六甲三宝山墓碑的整理,补充了17世纪至19世纪间早期华人移民的族群与商业信息,使区域华人史的研究基础更为坚实。

这些碑刻的类型与数量极为可观。马来西亚、印度尼西亚和泰国三国的这类寺庙碑文数量合计超过2500方,墓碑文超过1900方,会馆碑文接近300方,宗祠碑文超过200方,另有杂项碑刻80余方。总体而言,傅吾康团队在三地记录的碑刻点位超过1100处,总碑刻数超过5000方。这些材料不仅是重建东南亚华人移民历史的关键,更是理解华人社团组织、宗教体系、公益活动、商业网络与跨海联系的重要依据。

新加坡的碑刻研究在这一传统下进一步发展。2017年,笔者与许源泰合作出版《新加坡华文铭刻汇编(1819—1911)》,<sup>⑥</sup>系统收录了1200余方题刻,并对所有内容进行TEI标注,使其具备可检索、可比对、可分析的数字化能力。我们也正在编辑《新加坡华文铭刻汇编(1911—2015)》,预计同样将收录约1200方题刻。墓碑方面,目前已有超过1200方墓碑录入并完成TEI标注,未来将形成专门的华人墓葬数字数据库。

以碑刻资料为基础,可以进一步分析族群、行业、地缘组织与捐献行为的模式。新加坡的碑刻按地点可分为寺庙、会馆、宗族会堂、学校、医院等类型,以寺庙与会馆最为集中;按文献类型则包括石碑、牌匾、对联与其他建筑铭文,反映出华人在宗教、文化与公益活动中的广泛参与。

在马来西亚柔佛州的新调查表明,华人庙宇数量远超傅吾康时代所见。近年的系统性实地调查显示,柔佛境内华人庙宇数量已达九百余座,其中新山、峇株巴辖、麻坡、居銮等地最多,显示19世纪—20世纪以来大规模移民潮与商业开发对宗教建制的深刻影响。地方庙宇的功能不仅仅停留在宗教层面,它们还承担着社区组织、互助网络、经济协调乃至地方政治活动的角色。碑刻与档案为理解这些功能提供了具体证据。

除了南洋华人社团本身的碑刻研究之外,越南碑铭研究的数字化进展也为区域史

研究树立了重要范例。“越南研究”(Vietnamica)项目由法国远东学院与越南本地机构合作,旨在整理并分析约40000件拓片与约25000方石碑,其中三分之二以上记录了个人向村社捐赠土地或财物的行为。这些碑刻的内容集中于捐赠者在宗教体系中的地位建构,反映出越南传统社会的宗教经济结构。例如,捐赠者常被授予“后神”(Hậu Thần)或“后佛”(Hậu Phật)等头衔,并在村社祭祀中享受永续的纪念性祭祀。这一制度在17世纪—20世纪广泛存在,并在碑刻数量上呈现持续增长。

“越南研究”项目的重要贡献,一方面在于对碑刻的全面数字化,另一方面在于对文体与语言系统的深入分析。许多碑刻以越南口语与文言汉语的混合形式书写,体现当地对汉字体系的再造与本土化使用。这类文献长期被忽视,其重新整理对于理解越南语言史具有深远意义。此外,项目团队还从装饰、书法、刻工、采石场来源、皇家与民间制作体系等角度展开研究,使碑刻研究从文本学延伸至物质文化史与艺术史。

基于这些碑刻资料,学者们得以研究越南传统市场的发展。涉及“肆”“市”“廛”等概念的碑刻,共计约160方。它们分布在越南北部与中部多个省份,尤其在红河三角洲地区最为集中。研究发现,越南存在一种与宗教紧密相连的“宗教性市场”(如三宝市)。这些市场拥有独特的税收与组织结构,与普通市场之间时常因特权争议而引发诉讼。然而,诉讼的结果往往并不会削弱宗教性市场,反而将其特权扩展至普通市场,使之在宗教利益的框架下被重新吸纳。因此,越南的市场体系呈现出宗教性与经济性并行发展的特质,乡村与城市市场之间的复杂互动,也可通过碑刻资料来具体观察。

这一类碑刻为理解地方社会的经济组织提供了第一手材料,包括市场设立的原因、位置选择、开市制度以及与区域网络的关系。许多城市,如河内、富安等,最初都是由若干村落逐步聚合而成的城镇,其市场系统源自原有的乡村市场,并在城市化过程中不断扩展。越南周期性市场的制度亦反映出地方社会调节竞争、分配资源的机制,各类市场之间通过错开开市日期以维持区域内的商业平衡。

尽管“越南研究”项目呈现了越南碑刻研究的阶段性成果,其网站可提供碑文的准确地理坐标,是数字人文的宝贵示例,但其收录的华人碑文数量仍然有限,部分如河内福建会馆等地的重要碑刻尚未完全纳入,大量实地调查仍有待开展。

### 三、开启新阶段:数字人文发展的前沿与展望

在东南亚华人民间文献研究不断深化的同时,数字人文方法逐渐成为推动该领域前进的重要动力。随着人工智能技术的快速发展,OCR与自然语言处理(natural language processing, NLP)为处理成千上万页扫描文件、碑刻照片与手稿提供了新的可能。民间文献的内容往往包含多栏版式、手写体、苏州码子数字体系,或因纸张受损、印刷或拓本噪点等,使得传统OCR难以准确辨识。然而,通过高质量训练数据的累积,模型准确度得以显著提升。初步的未经训练的OCR准确率为六成左右,而在经过系统标注与

反复微调后,可提升至八成以上。

在实际操作中,研究者通常采取“人工参与的机器学习”方式处理文献。先由AI模型完成初步识别,再通过人工校正建立高质量标注数据。这些数据可被用于模型的持续训练,从而提升对不同文类、不同书写风格与不同时代文献的识别能力。特别是在处理账簿时,由于苏州码子数字体系的独特性、账簿页式布局的不固定性以及常见的草书手写,使得该类文本的OCR难度更高。因此,研究团队通常需要为账簿设计专门的识别模型,并对不同账簿格式构建高度定制化的训练数据。

OCR技术的突破为更进一步的自然语言处理奠定了基础。通过命名实体识别(named entity recognition, NER),研究者可以从大量非结构化文本中自动提取人名、地名、庙宇名称、商号、官职、年份、捐款金额、仪式活动等关键资讯。通过自动化处理,原本需耗时数月甚至数年的人工录入工作能够在短时间内完成,并且可进行大规模量化分析。此外,自动段落分类、主题模型与语义搜索等方法也可协助研究者从海量文献中探索新的研究线索,使民间文献不再只是静态文本,而成为可以运算、比较以及可视化的动态资料。

在档案利用方面,以巴达维亚公馆档案为例,荷兰莱顿大学图书馆所藏28卷公案簿涵盖了华人社群在殖民时期的行政、经济与组织运作,是研究早期侨社的核心文献。我们计划基于前期40页的初步训练集,持续扩展OCR模型,使其能够处理全部会议记录,并建立一个可识别公馆官职、商业网络、村属来源、航运资料与捐献记录的命名实体系统。账簿的处理将是下一阶段的重要任务,特别是涉及商号往来、收支状况、市场价格与跨海贸易活动的部分,可为理解华社经济结构提供前所未有的细节。

为了使这些文献与分析结果能够公开、共享与持续增殖,我们开发了“南洋地理信息系统”(Nanyang Sites GIS Data)数字平台,作为东南亚华人民间文献的开放获取地理数据库。该平台支持在免费GIS软件“QGIS”中加载多层资讯,并可连接扫描件、OCR文字、TEI标注文件与各类图片,令使用者能够通过地图界面直观地了解各地庙宇、会馆与华人社区的分布状况。平台收录的图层包括台湾地区宗教遗址、新马印老华人遗址、民丹岛寺庙调查,以及新加坡华人庙宇、会馆、坟墓区和克拉玛圣迹等,是目前南洋华人研究最全面的地理数据库之一。

这些图层不仅呈现出庙宇与会馆在空间上的分布规律,也揭示了不同历史阶段华人移民的路径与聚落模式。例如,新加坡的宗教与会馆分布,呈现出明显的族群与方言群结构;马来西亚柔佛的庙宇密度,反映了19世纪末至20世纪初的移民潮;印度尼西亚诸港口的华人遗址,则对应殖民时期商贸网络的发展。通过GIS分析,可以进一步探讨庙宇之间的分香关系、各社区的迁徙轨迹、商业网络的港口连接,以及宗教分布与经济活动之间的关联。

数字人文在这一研究中的另一个优势,是使语言学、文本学、材料学与历史学连结起来。例如,通过对碑刻字体、刻工技法、采石场来源的分析,可以重建碑刻制作的工艺

链与社会分工;通过对文本语言的比对,可以研究不同方言群体在海外的书写习惯与语言变化;而通过捐献网络的量化建模,则可以观察社区内的社会地位结构、经济实力分布与跨区域连接。

随着各类文献持续被收集与数字化,未来研究的潜力将更加广阔。我们规划建立东南亚华人领袖的传记数据库,以新加坡人物传记数据库(Singapore Biographical Database, SBDB)为基础,扩展至整个区域;也计划开发东南亚历史地理信息系统,从新加坡、马六甲、雅加达、槟城、会安、曼谷等主要港口与华人聚落开始,逐步形成区域性的历史地理平台;更长远的目标,则是构建一套涵盖碑刻、账簿、契约、族谱、宗教仪式记录等各类史料的“海外华人史料数字档案”,并通过开放获取模式,让全球研究者共同使用和参与。

总而言之,东南亚华人民间文献的研究正在进入一个由数字人文驱动的新阶段。在传统文献搜集与文本考订的基础上,AI技术、GIS工具与开放数据平台将使区域史研究不再局限于对个别文献的解读,而能够在宏观结构、跨区域机制与微观社群互动等不同层面展开综合分析。

在田野调查、系统整理与数字化技术的共同推动下,东南亚华人民间文献的研究已呈现出前所未有的广度和深度。随着各地社团、庙宇与家族逐渐将旧式档案数字化或淘汰,研究者必须抢在文献散佚之前,建立跨区域合作机制,使这些珍贵的历史资料得以保存、分析和传承。这些文献虽来自日常生活与地方组织,却在跨海移民史的框架中发挥着关键作用。它们不仅记录了宗教活动和社区内部的互动,更呈现跨区域商业与社会网络的动态结构,反映出“海洋中国”在区域交流中的持续影响。

数字人文方法的引入,使民间文献从静态档案转变为可分析的结构化资料。通过OCR、NER、主题分析、文本比对与GIS空间呈现,研究者可以从庞大的文献中自动提取出关键实体和关系,并将它们纳入可视化与网络分析体系,从而建立跨时间、跨区域与跨语言的综合研究框架。过去难以处理的手写账簿、苏州码子数字体系、模糊碑刻照片及损毁卷宗,如今都有可能通过模型训练与人工协作获得相对可靠的转写结果。随着训练数据的不断累积,模型的性能持续提升,并对不同文类形成更精准的识别能力。

这种资料的结构化与分析能力的提升,使得许多原本只能通过个案研究处理的史料,如今有机会进行系统性与量化的探讨。例如,宗教分香网络可以通过时间序列与空间分布重建其扩散路径,捐献碑刻记录可用于分析跨海商业资本的来源与方向,会馆档案能够揭示殖民地时期侨社的政治组织与经济调控,账簿中的日常交易则可还原出地方物价、行业结构与跨国贸易的细部运行。区域史研究因而从孤立的文献比对,迈向更具整体性与解释力的历史模型。

通过系统性收集、数字化处理和开放获取平台的建设,东南亚华人民间历史文献将能够以一种前所未有的方式被阅读、分析与比较。它们所呈现的跨海联系、商业网络、宗教体系与社区结构,将构成理解全球移民史与区域互动史的关键基础。在这个过程

中,数字人文为传统史学带来的不仅是技术层面的革新,更是视野与方法的转变,使我们能够从碎片化的资料中看见更广阔的历史结构,从地方的日常书写中重构宏观的海洋世界。

展望未来,东南亚华人民间历史文献的研究将更依赖跨国合作、跨语言团队与跨领域方法。随着更多文献陆续被发现、扫描与整理,预计将在数年内形成一个规模庞大、类型多元的数字史料库。为了让这些资料能够持久保存并为学界所用,不仅需要学者的持续投入,也需要地方社群、会馆、庙宇及文史团体的参与。文化记忆的保存与历史文献的传承,不只是学术界的责任,更是跨世代共同维护的公共事业。

---

注释:

- ①陈荆和:《十七、八世纪之安唐人街及其商业》,载《新亚学报》(香港)第3卷第1期(1957年)。
- ②饶宗颐(编撰):《星马华文碑刻系年(纪略)》,载《新加坡大学中文学会学报》1969年第10期,新加坡大学中文学会。
- ③陈荆和、陈育崧(编著):《新加坡华文碑铭集录》,香港中文大学出版部1970年版。
- ④庄钦永(编著):《马六甲、新加坡华文碑文辑录》,载《民族学研究所资料汇编》第12期,台北:“中央研究院”民族学研究所1998年版。
- ⑤傅吾康、陈铁凡(编):《马来西亚华文铭刻萃编》,吉隆坡:马来亚大学出版部1982年—1987年版;傅吾康(主编):《印度尼西亚华文铭刻汇编》,新加坡:南洋学会1988年—1997年版;傅吾康、刘丽芳(编):《泰国华文铭刻汇编》,台北:新文丰出版公司1998年版。
- ⑥丁荷生、许源泰(编):《新加坡华文铭刻汇编(1819—1911)》,桂林:广西师范大学出版社2017年版。

责任编辑:周 慧