

## “时代经纬：迈向新文科的数字人文”学术研讨会暨第三届中国数字人文大会会议综述

邱玉倩 / 南京大学地理与海洋科学学院

李梦月 / 南京大学地理与海洋科学学院

赵小萱 / 南京大学地理与海洋科学学院

南京大学人文社会科学高级研究院与中国索引学会主办的“时代经纬：迈向新文科的数字人文”学术研讨会暨第三届中国数字人文大会于11月6日至7日在云端线上成功召开。会议对新文科创新发展背景下数字人文的研究范式、理论体系与应用实践展开了全方位讨论，共同探讨“新文科”战略背景下的数字人文发展趋势与挑战。会议围绕“数字人文的新场景与新视野”“时空地图与人文世界”“历史遗产与档案记忆建设”和“人文研究数字化与数字技术人文化”四个方面，设置12个议题展开思想交流与分享，共同探讨“新文科”战略背景下的数字人文发展趋势与挑战。此次会议为国内外专家学者及数字人文爱好者提供了学术交流与思想碰撞的平台，展示了数字人文领域研究的新理念、新技术与新方法，既有利于专家学者了解学科前沿动态，同时又可以加强学科间交叉融合与探索，适应“新文科”战略背景下的数字人文学科教育及知识生产模式变革与人才培养要求。

### 一、数字人文的新场景与新视野

马费成教授以“当代人文社会科学研究的新场景与新视野”为主题，系统提出了数字人文研究领域的四个“新”，阐述了数字技术对数字人文研究在研究范式、研究视野、研究伦理以及研究体系上的机遇与挑战，以及将来数字人文研究需要关注的问题。罗家德教授以“社会科学理论介入大数据研究的几个新方向”

为主题,论述了数据介入社会科学带来的方法论变革,既包括对传统社会科学经典理论的改进和提升,也包括社会科学的解释模型与预测模型中数据理论的影响。周宪教授从数字人文与范式的转型出发,以数字人文文史研究为典型代表,深刻探讨了当下数字方法与数字工具对人文科学的知识生产方式和人文学科理念产生的深刻影响。

数字人文为人文研究带来了新方法与新手段,也开拓了人文研究的视野。赵宇翔《从信息素养到算法素养:数字人文视域下LIS研究的新议题》,选择数字人文研究作为场域,探讨LIS学者在这一方向下对于算法示能性和算法素养的研究机遇和相关议题。秦萍等《数字人文背景下的图书馆服务探索》,基于引文主题及知识图谱分析方法,从Web of Science源数据出发,通过对六个重点国家和地区在数字人文方面的科研产出、研究主题、高频词、共词网络等展开分析,梳理出国内外数字人文研究主题的发展脉络。张光伟《计算思维与数字人文和新文科建设》,通过分析典型的数字人文项目和研究实践中体现的计算思维的环节和流程,用计算思维的框架来分析数字人文研究的模式。王丽华等《数字人文的学科交叉特征——基于美国数字人文项目的分析》,以其开展的数字人文项目作为研究对象,采用网络调研、内容分析等方法分析了典型的数字人文项目的合作特征。梁兴堃《守正能否创新?论文新颖性、传统性对科学影响力的影响——以我国图情领域为例》,分析了新颖性和传统性对论文学科影响力之间的关系,研究结果凸显了我国图情领域守正创新的特点。

数字人文研究所体现出的超学科方法正在不断地打破学科和非学科的边界,将知识生产和传播变革带入了一个全新的“数字人文”传播时代。周葆华《智能传播研究的知识地图》,对智能传播研究的知识地图进行挖掘分析,探讨了智能传播研究的主要领域、核心概念以及主要知识基础来源。夏凡等《社交媒体之于数字人文的影响研究——以Twitter为例》,通过从用户和研究内容的角度分析得知,数字人文公共组织、学者个人和企业三个身份类型的用户使用社交媒体有着不同的行为和动机,社交媒体正在影响和改变着学术实践的方式。梁玉成《1948到1996年中国社会转型中的国家与社会——基于微观个体数据的计算》,揭示出国家的理性与非理性的边界,找出淹没社会群体的命运,并提出一种新的历史研究方法。王芳蕾等《基于信息熵的重大焦点事件传播涌现现象识别》,探索伴随焦点事件的传播,新闻与人的互动时涌现现象的识别,研究结果表明,重大事件的报道与被关注,有其数学规律。何兆阳等《基于相对熵的国家政治不稳定性的动态演化研究:以阿富汗为例》,以可用的海量新闻媒体数据为研究对象,开发了一种新的政治不稳定性度量方法对一般的社会科学(特别是政治科学)进行定量研究。

数字人文结合了信息技术和人文研究,需要学科基础理论、元数据、新媒体和技术环境的多元融合,也面临着传统学术界的质疑与挑战。肖鹏《一场没有凶手的谋杀:主流学术范式下的数字人文及其批判性思考》,强调要有意识地同当前的学术范式进行“搏斗”和“反抗”,要谨慎数字人文学术共同体的固化倾向,扩大数字人文的外延,要区别于传统的评价机制,促使数字人文本身得到转型和成长。林如诗等《高校图书馆员数字人文合作实践:基于CAG数据库的奥古斯丁学术研究史梳理》,以书目控制和文献循证为研究方法,开展馆员与人文学者之间的微型数字人文合作实践,进而总结出馆员与人文学者在资源受限下的全流程与多要素的数字人文合作机制。邓璐芾等《数字人文视角下老子思想研究的组织与推进》,结合数字人文理念、方法、技术及基础设施四个层面推进老子相关研究,并着重介绍了为老子思想研究构建的数字人文基础设施——老子思想专题研究平台。邱劼《博物馆数字藏品系统和数字策展人》,介绍了三个数字策展项目,并结合现有研究成果提出了“共同策展”的概念。

## 二、时空地图与人文世界

在跨学科知识范式下,数字人文的研究范畴随着不同学科体系的拓展而更为广泛。通过引入地理信息系统(GIS)技术,利用其空间数据管理、空间分析和可视化功能对历史空间进行展示及诠释,有效地支撑了历史学科的研究。韩昭庆《清初三大大图全图的数字化意义及案例分析》,讨论了对康熙《皇舆全览图》《雍正十排图》和《乾隆十三排图》进行数字化处理的科学意义及学术意义,同时分析了有关中国水系记载的历史文献的特点和传统,阐释康熙《皇舆全览图》在水系变迁研究中的重要性。沈婕《基于人文空间数据的叙事地图架构设计与可视化方法》,探讨如何从人文空间数据出发,构建、表达叙事地图的维度模型以及表达机制与方法,构建面向文化遗产保护的叙事地图架构并实现其可视化。潘威《基于TGIS与大数据技术的专项历史地名库的设计与搭建——以“数字历史黄河”地名库为例》,将时空框架和大数据技术引入历史地名数据库中,为历史地名信息系统的开发提供一种基于时空框架和时态地理信息系统(TGIS)的思路与方法,构建了“数字历史黄河”地名库,搭建了适用于黄河史研究的历史地名信息管理系统,提供了有针对性的时空分析工具和基于大数据的历史地名更新机制,为政区和聚落地名之外的水利地名等特殊历史地名的信息化管理提供了建设范式。陈刚《“图绘江苏”数字文化遗产建设——利用方志舆图重塑江苏历史文化图景》,以“图绘江苏”数字文化遗产建设为主题,利用中国国家

图书馆藏特色资源（方志丛书·江苏）电子方志库为基本文献来源，并参用哈佛燕京图书馆中国珍稀文献库等域外数字方志资源，运用数字人文、GIS、图像数据库等技术，探讨了“图绘江苏”数字文化遗产建设的项目内涵及学术价值，揭示了江苏方志舆图的时空分布与演变，及其所重塑的江苏历史文化图景。李文化《〈癸亥年更流部〉东南亚更路航线与地名的数字人文初探》，利用数字人文方法对东南亚更路地名进行考证，并得出比较可靠的结论，此研究对中国南海航海历史及郑和下西洋的研究有重要参考价值。

在面向数字人文的视域下，分析数字文化遗产的时空演变历程，探讨其空间分布格局、时空演变规律，对数字文化遗产的研究为历史地理研究和构建区域历史地理系统提供有力支撑。沈孙乐等《全生命周期下的明清官式建筑遗产知识的本体建模》，提出了一种建筑遗产知识的本体建模方式。针对明清官式建筑的本体建模，提出了新的赋予遗产信息互操作性的方式，有效地支持了明清官式建筑遗产领域的知识组织，也为进一步构建泛化的中国古建筑遗产本体提供了基础。杨璐嘉《中国私家藏书楼故址时空分布及其影响因素研究》，通过构建私家藏书楼的时空数据模型，展现了中国私家藏书楼分布中心自唐代到民国六个历史时期的时空演变历程，采用GIS空间分析技术、可视化以及传统地理模型等技术方法，探讨了中国私家藏书楼的空间分布格局，及其与自然环境、人文因素之间的互动关系，揭示了中国私家藏书的时空演化规律。杨谦雅等《数字人文视域下专题档案的空间、时间、主题多维度开发利用——以“乾隆南巡”专题档案为例》，提出时间、空间与主题维度下专题档案资源建设和数字人文技术相结合的开发利用实现路径，并以乾隆南巡档为实例，绘制时空地图，搭建“跟着乾隆去打卡”的专题网站。陶嘉洋《宋元以来闽南地区贸易海港与经济腹地的空间分析——以福建安溪为例》，基于相关历史文献和考古发掘成果，考证了古代安溪山川城邑地望，旨在为复原安溪境内古代交通路线作参考；利用地理信息系统（GIS），计算出古代安溪地区手工业作坊至闽南贸易海港理论成本最低的陆运路线，并窥探宋元以来作为经济腹地的安溪地区与闽南贸易海港之间的贸易路线。邱玉倩《基于空间句法的南京市鼓楼区空间形态演变研究》，以不同时期的数字化历史地图为研究对象，通过空间句法理论分析了南京鼓楼区的道路网络演变进程，进一步研究城市空间形态变迁与其使用功能的关联性。

### 三、历史遗产与档案记忆建设

数字人文在历史遗产与档案记忆建设领域受到广泛应用。大会围绕此部分议

题展开讨论,其中又可以细分为“艺术与遗产的将来”“数字历史叙述的有效与有限”“档案与记忆的数字重塑”三方面。

数字人文为艺术与遗产的知识管理与文化振兴提供了新的方法与实现路径,是艺术遗产保护的重要途径。陈静等《基于CascadeRCNN检测模型的传统纹样识别算法研究》,针对传统纹样图像识别,以CascadeRCNN作为纹样检测基础模型进行算法改进,提升了模型对旋转目标的检测能力、骨干网络特征表达能力以及模型检测框的准确性。朱子彤《吕班影片研究及其特点:基于数字人文工具的考察》,通过吕班拍摄的电影和其相关联的文献进行可视化分析,找寻吕班的电影风格分析特点与电影图像本身结构的联系,为电影研究提供新的尝试路径。秦翼《数据深层标注与影史微观研究——一次电影史量化研究的尝试》,以上海孤岛时期《申报》图片式电影广告进行了数据标注与统计观察,使图片式的电影广告史料变为可统计可观察的数据并展开分析。高明《Kubikat书目检索系统的关键词体系:专业性图书馆如何建立研究框架》,从讨论Kubikat书目检索系统中的关键词体系出发,探讨如何在中国利用Kubikat中的德语关键词和中国经、史、子、集的图书分类系统来建立中国艺术史图书馆。李梦琦等《非物质文化遗产传统技艺知识本体构建——以锦、绣、年画三类为例》,针对当前非物质文化遗产中传统技艺类专业知识等问题,构建了一种结合专家经验和文本分析算法结果的知识本体,实现对于不同来源、不同种类的非遗核心知识的体系化构建、结构化组织和语义化描述。包艺等《数字人文视角下南阳汉画乐舞图像研究》,基于“中国汉代图像信息数据库”,对画像石画面描述的术语文本、墓葬信息及断代年份的统计与量化分析,从宏观层面探索汉画乐舞图像在墓葬场域中的独特意义及汉画图像元素之间的相关关系。万翔等《计算机辅助岩画分析:网格化与语义化方法》,将大地测量学中的经纬坐标体系迁移运用到对岩画的计算机辅助分析之上来准确描述岩画物理形态、画面元素及元素组合情况,以一种相对客观、科学的角度认知并解析岩画内容及意义。赵珂《数字人文背景下文化遗产可视化的发展——走进数字叙事可视化空间》,基于可视化和叙事学的理论,以“邂逅春天——数字江南文化”项目为案例,探索DSV在博物馆CH中的应用前景。张轩慧等《基于因果推断的数字人文文化遗产众包平台分析》,从任务特征的角度研究文化遗产众包(CHC)的志愿者参与不平衡问题,提出了原则性的定性因果推理方法与缓解问题的设计框架。

会议中针对“数字人文”视野下历史学科叙述的有效与有限问题引发激烈讨论。赵小萱《面向〈拉贝日记〉文本的地理及事件信息挖掘与组织》,以《拉



贝日记》为例,提出了一套完整的从历史文本中挖掘地理实体及其相关信息并可视化展示的方法与流程。王林旭等《基于OWL-Time的时间本体在古籍中的应用》,提出了一个基于通用时间本体OWL-Time的面向古代中国历法的中国时间本体——中国传统时间本体(ACTO),提供了中国传统的表示方法与格里高利历方法之间的对应关系,将时间信息的文字描述翻译成可计算的通用标准时间戳。邹亚文《大数据时代的数字科学史面临的机遇和挑战》,从计算工具的使用、跨学科领域的工作模式、数字项目的保存和新时期历史学家的培训四个方面详细讨论了历史学家面临的更深层次的挑战。蒋杰《我们需要什么样的数字历史:经验与反思》,立足既有经验对数字历史研究的定义问题、性质问题和应用问题展开反思和讨论。吴静《哲学与史学的对话:算法是否会改写历史书写方式?》,认为数字技术不仅为人文研究提供技术工具支撑,数字权力和数字技术逻辑更凝聚着历史记忆。战蓓蓓《史学、科技与应用——数字人文视野下学科交融碰撞与历史学本位思考》,从自身项目经验出发,对历史学科为本位的方法论、学科的交叉与区分、历史与应用、历史与大众、历史与现代的关系进行了反思。

随着数字人文在档案学与数字记忆领域的不断深入,新的档案资源建设与记忆重塑路径亟需发展。加小双等《数字人文赋能档案资源体系建设:机制与路径》,集中探究数字人文赋能档案资源体系建设的作用机制与发展路径,提出数字人文赋能档案资源体系建设的三维机制体系。李梦月《历史建筑文化地图的叙事性建构》,聚焦历史建筑遗产的非空间化保护,结合文学叙事与地图学相关理论提出“时空切片—叙事结构—叙事评价”的叙事性构建的理论与方法。牛力等《融合、建构与服务:数字记忆视角下的学术名人知识库研究——以吴宝康学术名人知识库为例》,形成了传统叙事之外的学术名人档案知识化复合研究路径。姜婷婷等《数字记忆空间理论探索》,重视“人”在记忆保护与传承中的主观能动性,围绕“人—数字化—记忆”三个关键维度进行理论梳理,最终整合形成“数字记忆空间”的理论框架、构建模式与发展方向。邓君《近代报纸文献的重组再构思索》,从图情学科知识组织视角出发,就近代报刊的重组再构展开思考,以推动近代报纸文献的史料价值、学术价值和历史史实的订正价值开发、利用与共享。侯西龙等《文化遗产关联数据聚合与可视化分析》,针对文化遗产图像的资源,提出多源异构图像关联数据聚合与语义组织的方法,利用关联数据、知识图谱、IIIF等技术建设开发了“海外博物馆里的中国——文物数字图像关联数据聚合平台”,探索对海量文化遗产图像资源的知识挖掘、可视化分析与叙事重构。

#### 四、人文研究数字化与数字技术人文化

数字人文在文学与文本中的拓展与应用主要围绕“人文研究数字化”与“数字技术人文化”两个方面展开。大会设立了“数字人文视野下的情感、话语与思想”“文学与数字人文”和“数字人文视野下的植物与空间”三个议题，旨在探讨数字人文技术在情感识别、翻译文风、计量计算等领域的应用，以及在文学文本研究中应用知识图谱、命名实体识别等数字技术。

情感、话语与思想是文本研究探讨的重要议题。黄紫荆《基于Transformer的近代名人日记情感识别与可视化研究——以〈拉贝日记〉为例》，建立了一种考虑远程上下文、具体语境的情感分类模型，结合时序分析、地图映射可视化拉贝的个人情感画像。胡悦“Refocusing Democracy: the Chinese Government's Framing Strategy in Political Language”，以计算机辅助文本分析，聚焦分析《人民日报》50年来100多万篇政治文章，得出结论：中国政府以国家治理和发展的主要任务定义民主内涵，并积极与既有民主话语形成联系，保持对话。刘凯然《电子游戏与被计算的历史：〈十字军之王〉中的拜占庭历史叙事》，提出目前电子游戏作为数字媒介，已经广泛被用作历史叙事载体，所以电子游戏也应被纳入“数字史学”的研究范畴的观点。苗威等《“道”与“器”之间：基于古史体系建构的数字人文实践》，在东北边疆历史问题研究实践验证“数字人文”的综合交叉特性，符合新文科发展趋势，呼唤在历史研究中以数字技术为“器”，追求客观历史之“道”。左家莉《数字人文视野下的数据偏差问题》指出，虽然蓬勃发展的数字人文项目建立了坚实的数据基础，但是在应用人工智能与自然语言处理技术时，数字人文学者需要关注和警惕数据“偏差”（bias）问题。

数字人文技术拓展文学研究范畴，为文学研究提供新思路。王峰《文学计算应该成为独立学科》，从团队的技术实践经验出发，认为社会计算是数字人文当前非常有价值的一个部分，研究应该从单纯的工具化转向专业的工具化。赵薇等《作为“新文类”实践的“散文诗”——基于多特征建模与网络分析视角的再认识》，为研究现代散文诗的发生与形成，采用多特征建模、文本聚类、机器学习和网络分析等方法，搜集整理，并聚焦于民国时期出版的上千篇散文诗译介和创作文本，旨在探索新文学奠基期文类惯习定型的普遍规律及其和世界文学的关系，构建一个较全面的文学场图景。诸雨辰等《中国古代文体相似度的计量研究》，尝试以定量计算的方法，通过对文本的形式特征进行统计计算，计算不同文体之间的相似程度并尝试计算机的自动聚类。旨在重新审视中国古代的文体

分类及探究不同文体之间的相互关系。刘洋《科幻与推理小说中的解谜叙事——基于词频动力学的远读与可视化研究》，通过分析《三体2》与《东方快车谋杀案》的关联词频时序图，直观体现文本叙事主线、逻辑推演、故事高潮、解谜收束等结构性特征。通过分析大量文本的平均词频集聚函数，发现解谜峰在推理小说组别的图像上是一个最为显著的特征，这表明解谜型叙事在推理小说创作中被广泛使用。李山雨等《基于词向量主题模型的小说人物话语主题分析——以〈水浒传〉为例》，基于词向量主题模型（ETM, embedded topic model）对《水浒传》任务话语内容进行主题分析，旨在考察人物话语中所蕴含的主体结构以及分析人物话语主题的人相对远近关系及其随时间演化的过程，揭示人物心理的前后转变及相似任务群的变化。郝若辰等《从永明到景龙：初盛唐近体诗声律规则的还原与重构》，文章参照多样的衡量语料选择与声律统计的标准，使用清华大学开发的“古典诗歌声律分析系统”，在“元兢—崔融声病”与平仄律两个规则体系中，对初盛唐作品分诗体、分时段、分创作群体进行统计，得到从永明到景龙时期从四声系统到平仄律的演进脉络。孙燕等《〈春秋左传正义〉引书知识的标注与计量》，基于《春秋左传正义》，采取正则表达式方式匹配出4,421条典籍内容引用及4,129条作者内容引用，采用Neo4j构建引书知识图谱，并且量化分析引书计量、多角度对比分析注引及疏引、探讨作者关系及流派传承。借用数字人文技术手段，从新视角解读《春秋左传正义》。李佳纯等《〈论语〉在儒家典籍中的文本复用计量研究》，基于近似最长公共子串匹配方法，识别和抽取上始先秦下迄近代的136本儒学经典书目中复用《论语》的文本单元，并从宏观的计量视角对《论语》篇章、条目、短语三个维度的复用情况进行可视化统计分析。柳青秀等《川端康成〈雪国〉中文译本的文风分析》，使用计量方法对比川端康成《雪国》原文与两版不同译本及两位译者原创作品的文风，发现两版译本文风类似，并区别于两位译者原创作品的文风，得出《雪国》一书的文风可以被翻译的结论。郝雪丽等《敦煌诗歌写本群知识图谱构建研究》，基于科学计量研究方法，以敦煌诗歌写本群为研究对象，通过文本细读，对写本形态、内容进行写本学分析，梳理出有内容重出关系的诗歌写本群六十余个，涉及敦煌诗歌写本一百余件，并探讨敦煌诗歌写本群知识图谱数据库建设与智慧服务的开发，从数字人文视角，立体呈现敦煌诗歌写本间的关系。

数字人文技术为古籍内容处理与特定信息抽取提供新方案。马开颜等《数字人文视域下的植物文学内涵研究》，基于多树种的文本语料，采用LDA模型，对比不同树种的文学内涵，探究植物与文学内容隐性联系以及各类植物文学内涵的共性与个性。李惠等《基于〈方志物产·云南卷〉的物产名智能考证



初探》，针对考辨物产“同名异物”与“同物异名”的问题，利用深度学习、文本分析、社会网络分析等领域的原理和方法，为考证物产提供新思路。吴梦成等《数字人文视角下的先秦典籍植物命名实体自动识别研究》，利用多种预训练语言模型构建基于先秦典籍的植物命名实体自动识别模型，最终选用SikuBERT和SikuRoBRETa模型，作为古汉语典籍植物命名实体自动识别的最佳模型。王东波《Sikuaip：面向数字人文的中国古代典籍智能处理平台》，介绍了免费中国古代典籍文本智能处理平台Sikuaip，平台使用古籍语料训练的SikuBERT和SikuRoBRETa模型，提供自动分词、自动分类、词性标注、实体识别、自动断句和标点功能，旨在为数字人文研究者提供高质量的文本处理服务。程齐凯等《WHUIMS翻译引擎的构造与应用》，基于词对齐的同义数据扩充和基于规则的模板数据补充等方式，构建了“文言文—现代文”翻译数据集，并通过Transformer训练得到WHUIMS文言文翻译引擎。经过与百度翻译和微软翻译对比，WHUIMS引擎能够实现文言文文本的快速和高精度翻译。

## 结 语

在全球新科技革命、新经济发展、中国特色社会主义进入新时代的背景下，传统文科亟待突破传统思维模式，以继承创新、交叉融合、协同共享为主要途径，促进多学科交叉融合，以期迈向新文科。数字人文作为数字技术与人文学科交叉融合形成的新兴领域，是传统文科迈向新文科的重要途径之一。本次会议除学术汇报与海报发表外，为鼓励青年学者能在数字人文学术共同体中得到更好的建议与发展，会议组还特别策划了“青年沙龙”，由北京大学中文系博士生高树伟与南开大学文学院讲师张昊苏担任召集人。此次年会，汇聚众多学科的专家学者，报告内容包罗万象，多学科领域的内容碰撞发展，形成热烈的讨论，共同畅想新文科的建设与发展。本次年会为国内外的数字人文专家学者与爱好者提供了思想碰撞与展示成果的良好平台，这不仅有利于开拓视野，宣传最新的数字人文成果，更推动了数字人文的研究发展，加快了向新文科迈进的步伐。

（编辑：李惠）