



Research Center for
Digital Humanities of
PKU
北京大学数字人文研究中心

2022

北京大学数字人文研究中心 工作进展

Achievement of the Digital Humanities Research Center @ PKU



北京大学数字人文研究中心工作进展

Achievements of the Digital Humanities Research Center @ PKU



2022-3-30

北京大学数字人文研究中心

<http://DigitalHumanities.pku.edu.cn>

目 录

一、北京大学数字人文研究中心简介.....	2
二、北京大学数字人文研究中心大事记.....	4
三、北京大学数字人文研究中心进展.....	5
1. 学术活动.....	5
1.1 “数字人文视角下的典籍传承与学科融合”论坛.....	6
1.2 “学科电子资源+数字人文资源”导航讲座.....	7
1.3 数字人文云端国际工作坊.....	7
1.4 “数字人文视角下的中国历史研究”国际云端研讨会.....	8
1.5 “智能时代古籍数字化资源的开发与利用”研讨会.....	9
1.6 北京大学首届数字人文作品展.....	10
1.7 “古籍智能信息处理”系列研讨会.....	11
2. 跨学科合作、校企合作与国际合作.....	12
2.1 与哈佛大学的全面国际合作.....	12
2.2 字节跳动公益捐赠古籍数字化研究.....	13
2.3 校内跨学科的交流与研讨.....	14
3. 课题研究.....	17
“国家古籍数字化工程”专项经费资助项目	17
国家自科基金国际重点项目“儒家学术史知识图谱构建研究” ..	18
中宣部出版局委托项目“古籍数字化技术创新与应用” ..	18
中美合作“中国历史人物资料库”(CBDB)在线查询系统	18
4. 论文发表.....	18

一、北京大学数字人文研究中心简介

“北京大学数字人文研究中心”(PKUDH)是北京大学的校级跨学科研究虚体。中心于2020年5月经北京大学批准成立；2022年2月在北京大学人工智能研究院下设立数字人文研究中心。中心是中国古籍保护协会“古籍智能”专业委员会的牵头组织单位，致力于古籍资源的智能开发与利用。2022年3月接受字节跳动公益的捐赠，以“北京大学-字节跳动数字人文开放实验室”为研发基地，将中心在古籍智能信息处理的技术积累转化为面向大众的公益性社会服务。

数字人文有两方面的含义：大数据和人工智能技术使得在长历史跨度和海量文献资料上的研究成为可能，为人文学科提供了新方法、新工具，催生新范式；另一方面，泛在的数字信息环境更需要人文精神的引领，数字人文是数字与人文相沟通的通道。数字人文要求全科型的人才培养：新一代的人文学者既要具备深厚的人文素养，又要掌握计算思维与数据驱动的研究方法；理工科的学生既要接受系统的科学训练，又要泽被人文精神的熏陶，以此培养智能时代对人类社会和人类命运有整体性思考和把握能力的全科通才。

北大数字人文中心的宗旨是打造数字环境下人文学科发展的信息基础设施，探索智能时代数据驱动的人文研究范式。中心为人文社科学者与理工科专家提供交流与协作渠道，一方面推动大数据、智能技术在人文社科领域的应用，另一方面倡导技术创新要树立以人为本的大局观，促进信息技术与人文社科的双向融合发展。北京大学数字人文中心在校内致力于组织跨学科的研究力量，广泛吸引各院系师生参与，建构协同创新的数字人文平台；在校外积极与国内外高校和相关学术机构建立广泛的学术合作，筹办高端学术活动，培养全科型人才，推动跨学科、跨地域、跨文化的数字人文研究项目。中心目前的组织结构如下：

■ 学术委员会

邓小南（主任），北京大学历史学系教授

包弼德（Peter K. Bol），哈佛大学东亚语言与文明系讲座教授

李晓明，北京大学计算机科学技术系教授

朱青生，北京大学历史学系教授

荣新江，北京大学历史学系教授

刘玉才，北京大学中国语言文学系教授

赵白生，北京大学外国语学院世界文学研究所教授

罗新，北京大学历史学系教授

李文新, 北京大学计算机科学技术系教授

■ 工作委员会

王军 (主任), 北京大学信息管理系教授

史睿, 北京大学中国古代史研究中心副研究馆员

苏祺, 北京大学外国语言学及应用语言学研究所副教授

杨浩, 原北京大学哲学系、儒藏编撰中心助理教授

■ 专家委员会

陈斌, 北京大学地球与空间科学学院教授

程乐松, 北京大学哲学系教授

何晋, 北京大学历史学系教授

赫忠慧, 北京大学体育教研部教授

李道新, 北京大学艺术学院教授

李咏梅, 北京大学工学院研究员

刘琳琳, 北京大学外国语学院日语系副教授

罗定生, 北京大学智能科学系副教授

沙志利, 北京大学《儒藏》编纂与研究中心副研究员

邵燕君, 北京大学中文系长聘副教授

宋亚云, 北京大学中文系长聘副教授

童云海, 北京大学智能科学系教授, 图书馆副馆长

王长松, 北京大学城市与环境学院副研究员

王腾蛟, 北京大学计算机科学技术系教授

王彦晶, 北京大学哲学系长聘副教授

杨海铮, 北京大学中文系教授

杨梦, 北京大学外国语学院世界文学研究所助理教授

杨韶蓉, 北京大学《儒藏》编纂与研究中心副研究员

叶炜, 北京大学历史学系教授

俞敬松, 北京大学软件与微电子学院副教授

张剑嵒, 北京大学考古文博学院助理教授

张鹏翼, 北京大学信息管理系长聘副教授

赵丹群, 北京大学信息管理系教授

二、北京大学数字人文研究中心大事记

- 2022 年 3 月 17 日 北京大学首届数字人文作品展开幕
- 2022 年 3 月 17 日 接受字节跳动公益的捐赠，以“北京大学-字节跳动数字人文开放实验室”为基地，促进古籍资源的智能开发与利用
- 2022 年 3 月 12 日 在高校古委会和中国古籍保护协会的支持下，由数字人文中心筹备的“古籍智能系列在线研讨会”正式启动
- 2022 年 1 月 13 日 在北京大学文化产业研究院燕南园 52 号举办“文化遗产与文化计算”学术沙龙
- 2021 年 12 月 30 日 在北京大学校长办公会上通过为数字人文研究中心配置基本资源的决议
- 2021 年 11 月 24 日 2021 年度国家古籍数字化工程专项资助入选项目名单公布，北大数字人文中心主持或参与的项目达 5 项，占全部资助项目总数的 18%。
- 2021 年 9 月 15 日 作为牵头单位，联络高校、图书馆、企业等相关单位，发起成立中国古籍保护协会古籍智能开发与利用专业委员会
- 2021 年 8 月 10 日 数字人文研究专题工作会议在燕园大厦 1211 会议室举行，通过促进数字人文中心实体化建设的决议
- 2021 年 6 月 16 日 在北京大学行政月度会上汇报“北京大学数字人文发展规划倡议”
- 2021 年 5 月 16 日 在北京大学校长办公会上汇报北京大学数字人文进展及发展规划
- 2021 年 1 月 北京大学数字人文中心-哈佛大学费正清中心全面合作协议签署
- 2020 年 12 月 国家自然科学基金国际重点合作项目“中国儒家学术史知识图谱构建研究”获批
- 2020 年 9 月 25 日 “智能环境下古籍资源开发与利用专题研讨会”顺利召开
- 2020 年 6 月 6 日 “数字人文视角下的中国历史研究”云端国际论坛举办
- 2020 年 5 月 27 日 “北京大学数字人文研究中心”正式宣布成立
- 2020 年 4 月 18 日 北京大学“数字人文国际云端工作坊”成功举办
- 2020 年 3 月 20 日 “人文学科电子资源与数字人文导航”系列在线讲座
- 2020 年 1 月 16 日 北京大学数字人文建设与发展研讨会
- 2019 年 12 月 17 日 北京大学-哈佛大学数字人文合作研讨会
- 2019 年 10 月 13 日 “数字人文视角下的典籍传承与学科融合”研讨会

三、北京大学数字人文研究中心进展

北京大学数字人文研究中心自 2020 年 5 月正式成立以来，各项工作推进迅速。中心以古籍资源的智能开发与利用为核心研究目标，在国家自然科学基金、国家古籍数字化工程专项资助等国家课题的支撑下，在古文自然语言处理的核心技术上取得突破性进展。在校内通过系列学术活动凝聚北大数字人文的跨学科研究力量，培养师生们在数字人文领域探索的兴趣和技能；在学术界将古籍领域和信息技术领域的专家联系在一起，组建中国古籍保护协会古籍智能专业委员会，筹办全国范围内的古籍智能系列研讨会。在国际合作方面，中心与哈佛费正清中国研究中心签署全面合作协议。2022 年 3 月北京大学首届数字人文作品展的举办将北大数字人文的发展推向了高潮，标志着数字人文事业赢得了全校各专业、各层面的普遍支持，为未来发展奠定了坚实基础。在 3 月 17 日作品展的开幕式上，字节跳动公益宣布向北大数字人文中心捐赠基金，并调动企业技术团队与数字人文中心共同打造古籍智能化整理平台和古籍开放阅读平台，向全社会提供公益服务，构建古籍文化普及传播的全媒体渠道。经过上述多个层面的努力，北京大学数字人文研究中心确立了在古籍智能信息处理领域的领军地位。下面，按照时间顺序从学术活动、合作联系、课题研究、成果发表四个方面来概述中心取得的成绩。

1. 学术活动

2019 年，在北京大学文研院举办了“数字人文视角下的典籍传承与学科融合”研讨会，对北大在数字人文领域现有的成果积累进行了总结。

2020 年受疫情影响，高校无法展开正常的线下教学活动。在此情形下，北京大学数字人文中心率先举办了一系列在线学术活动，带动了我校课堂教学之外的线上学术活动。自 2020 年 3 月中旬开始到 6 月初，数字人文中心先后共组织了三场大型的学术活动，面向全校师生推广和普及数字人文的理念、方法与工具，在数字人文领域造成了广泛的影响。此外，中心受中宣部出版局全国古籍管理规划办公室委托，于 9 月底邀请了国内从事古籍整理、出版与研究的十三家单位的 40 多位专家学者，举办古籍数字化专题研讨会，为国家十四五期间的古籍数字化做预研。

进入 2022 年，数字人文中心举办的活动更加丰富，影响范围也更广。在高校古委会和古保协会的支持下，由数字人文中心发起并筹备的“古籍智能信息处理”系列研讨会正式启动。2022 年 3 月 17 日，北京大学首届数字人文作品展正式开幕，将北大数字人文发展推向了高潮。各项活动简介如下。

1.1 “数字人文视角下的典籍传承与学科融合”论坛



2019年10月13日，由北京大学人文社会科学院、北京大学信息管理系数字图书馆实验室、北京大学中国古代史研究中心和商务印书馆联合主办的“数字人文视角下的典籍传承与学科融合”论坛在北京大学静园二院208室召开。北京大学副校长王博、北京大学人文社会科学院院长邓小南、中宣部出版局古籍处一级调研员章隆江、北京大学信息管理系主任张久珍到会并致辞。来自北京大学、清华大学、中国人民大学、复旦大学、武汉大学、南京大学、浙江大学、北京师范大学、湖南大学等高校的数字人文资深研究人员，以及全国古籍整理出版规划领导小组、国家古籍保护中心、中华书局等相关机构的约120余位专家、学者出席了本次论坛。

在本次论坛上，北京大学信息管理系王军教授详细介绍了他带领数字图书馆实验室的同学们用了一年时间所完成了“《宋元学案》知识图谱可视化系统”。该系统将240万字《宋元学案》所记录的两千多位宋元理学学者、近100个学术流派所涉及的人物、时间、地点、著作等提取出来，构造成知识图谱，提供可视化展现、交互式浏览、语义化查询等功能。

参加论坛的各高校代表也纷纷介绍了各高校在数字人文领域现有的成果积累，研讨数字人文的发展。北京大学中古史研究所研究馆员史睿主持了随后的“数字人文与学科融合”的圆桌讨论，来自学界和业界的二十余位专家学者从各自研究的角度阐述了对主题的见解和看法，展开了热烈的讨论。与会嘉宾对北京大学在沉静幽深的静园二院举办这样一场前沿领域的热烈研讨会表示感谢，希望北大能继续发挥在人文社科领域的引领作用，与兄弟院校一道共同推动数字人文在中国的发展。

1.2 “学科电子资源+数字人文资源”导航讲座

疫情期间师生们在家中办公与学习，为响应师生们利用电子资源的迫切需求，筹建中的数字人文研究中心与北京大学图书馆通力合作，率先组织大规模的线上学术活动，于 2020 年 3 月 20 日推出“人文学科电子资源与数字人文资源导航”系列网络直播讲座。本次系列讲座面向中文系、哲学系、历史学系和考古文博学院的师生分别举办四场，于 3 月 29 日结束。师生们踊跃报名参加，四场讲座的人数从 150~180 人不等。参加线上讲座的不仅有我校的师生，还有外校甚至国外的学者和学生。每场讲座两个小时，分成两个部分：前一个小时由图书馆信息素质教育团队讲解该学术领域常用数据库的检索与利用，后一个小时由数字人文实验室的成员介绍数字人文的常用研究方法、工具和典型案例。配合讲座，数字人文实验室专门开发了“数字人文资源导航”网站，汇总全球的优秀数字人文网络资源。

1.3 数字人文云端国际工作坊

在成功举办“数字人文资源导航”讲座的基础上，数字人文中心随后在 2020 年 4 月 18 日至 5 月 23 日期间组织了“数字人文工作坊”，进一步培训数字人文的常用工具和研究方法。工作坊有三个模块构成：学者授课、学员分享和专题研讨。

工作坊邀请了德国马克斯-普朗克研究所、德国柏林国家图书馆、台湾大学和北京大学共六位领域专家讲授数字人文工具与应用实践，包括：中文领域三大数字人文研究平台 DocuSky、MARKUS、LoGaRT 的使用方法，利用 R 语言和 GIS 地理信息系统展开数字人文领域的研究等。共有 379 份海内外众多院校师生填写了报名问卷，其中有 180 人来自北京大学，199 人来自清华大学、中国人民大学、山东大学、南京大学等一百余所高校或科研机构，还有 15 人来自美国、英国、新加坡的海外高校。工作坊全程采取腾讯会议在线直播方式，累计在线听课达 1500 人次，1000 余人次观看课程录像回放。在线课堂氛围很好，学员踊跃提问，教师与学员之间、学员与学员之间互动积极，达到了很好的教学效果。在专家讲授之外，工作坊还组织学员分享课，三位来自北京大学历史系、信息管理系和香港大学人文学院的博士生分享了自己的数字人文技能。北大中文系邵燕君老师组织了题为“数字与文学的对话——数字人文范式对传统文学研究方法的挑战”的专题在线研讨会，邀请了数字人文领域的三位新锐学者姜文涛（浙江大学）、邱伟云（山东大学）、赵薇（社科院）分别讲述他们对数字人文环境下文学研究的看法，并对数字人文的源流变迁、学理定位、实际应用等进行了细致的分析。

和热烈的讨论。

通过工作坊，一方面向广大师生普及了数字人文的常用工具和研究方法，另一方面，在数字人文中心汇聚了感兴趣的师生和研究团队，形成一个数字人文的松散学术团体，有效推动了我校的数字人文发展。

1.4 “数字人文视角下的中国历史研究”国际云端研讨会



2020年6月6日，北京大学数字人文研究中心、人文社会科学院、历史学系和北京论坛联合主办“数字人文视角下的中国历史研究”云端国际论坛。研讨会以“数字人文视角下的中国历史研究”为题，邀请了在中国历史研究领域卓有建树的学者大家：北京大学历史学系邓小南教授，著名中国历史研究学者、哈佛大学东亚语言与文明系讲席教授包弼德（Peter K. Bol），长期致力于中国古代科学技术史研究的德国柏林马克斯·普朗克科学史研究所所长薛凤（Dagmar Schäfer）教授，比利时汉学家、荷兰莱顿大学区域研究所中国史教授魏希德（Hilde De Weerdt），以及台湾中研院历史语言研究所的陈熙远研究员，围绕在数字人文方法如何推动历史学发展、传统史学的格局将如何改变、研究者如何充分利用信息技术工具来从事研究、数字人文方法面临的挑战与需要突破的瓶颈等前沿学术问题，展开了深入的交流和对话。

此次线上论坛，通过北大宣传部、北大新闻网、北大官方微博，以及北大的海外宣传渠道都做了预热宣传。论坛通过哔哩哔哩网、百度直播网、中国网三大网络平台直播，直播人气数峰值达25万，成为2020年度国内数字人文界的一件

盛事。此次云端论坛汇聚了汉学领域的一众国际著名学者，造成了广泛而深远的社会影响，将上半年的数字人文活动推向了高潮。

1.5 “智能时代古籍数字化资源的开发与利用”研讨会



为了促进中华典籍在智能时代的深度开发与广泛利用，中宣部出版局于 2020 年启动了古籍数字化十三项调研课题，在全国范围委托了十三家机构展开专项调研。受中宣部出版局古籍处委托，北京大学数字人文中心邀请了上述承担课题的十三家单位和相关领域的 40 多位专家学者，于 2020 年 9 月 25 日在北京大学文研院举办了专题研讨会。研讨会全面汇总了数字环境下中华古籍的数字化开发与利用现状，并为智能环境下古籍数字化资源的开发与利用勾画蓝图，议定实现路径。

北京大学数字人文研究中心承担了“古籍数字化关键技术创新与应用研究”子课题，在此次研讨会做了三项专题报告：当前古籍数字化领域的关键技术的最近进展、古籍数字化整理平台和数字人文研究平台的发展现状、智能信息环境下数字化古籍文献资源开发与利用的整体解决方案。本次会议确立了北京大学在古籍数字化开发与利用领域的核心地位，为争取古籍数字化的国家重大工程项目做了铺垫。

1.6 北京大学首届数字人文作品展



2022年3月17日上午，北京大学首届数字人文作品展开幕式在北大全球大学生创新创业中心举行。第十二届全国政协副主席、中国科学技术协会名誉主席、北大前沿交叉学科研究院院长韩启德院士，北大校长郝平、副校长王博，北大燕京学堂名誉院长袁明教授，国家图书馆副馆长张志清，北京字节跳动副总裁李涛等嘉宾出席开幕式。开幕式由北大社会科学部部长强世功主持。

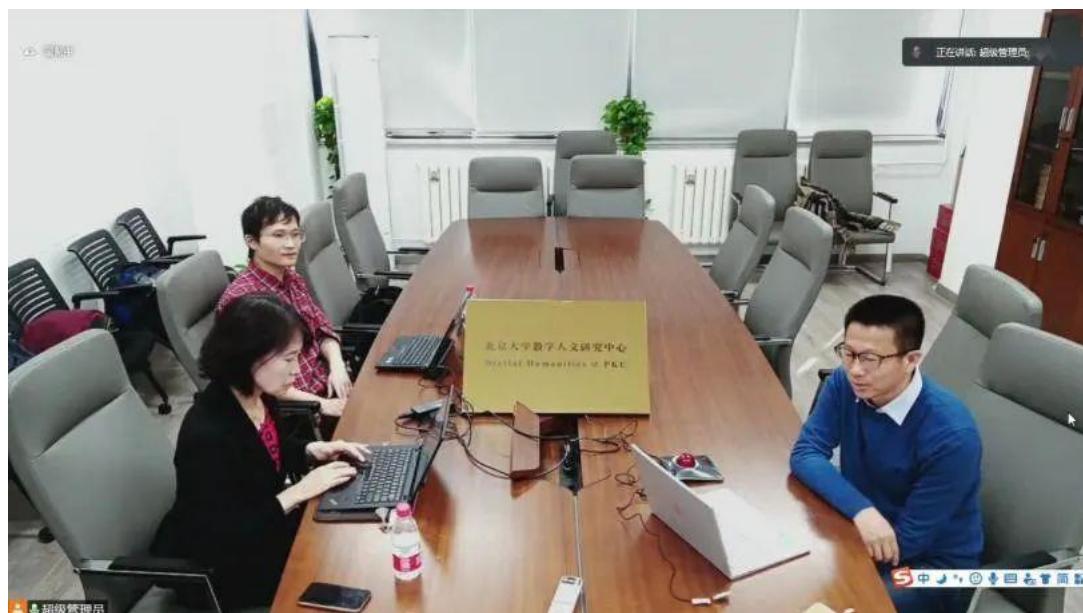
王博代表北京大学致辞，北大很多院系、老师都有与数字人文相关的丰富积累，像本次展览这样一个跨学科的学术活动展示了北大深厚的人文和科学积淀，也为北大几十年来推动的学科交叉和交叉学科发展开拓了一个新的领域。他希望北大人文学者积极面对数字时代的到来，理解数字人文本身的意义，开拓新的研究方式和新的研究领域。张志清表示很高兴回到母校，并为母校以数字人文为抓手发展新文科体系、开创“守正出新”的局面而赞叹。韩启德在致辞中表示，本次展览的各类作品异彩纷呈，展览涉及的主题很全面，内容专业性强，彰显了北大数字人文的实力和发展的决心。

随后，王博代表学校向李涛授予捐赠感谢状，感谢北京字节跳动对北京大学数字人文领域发展的支持。开幕仪式后，北大视觉与图像研究中心主任、历史学系教授朱青生，清华大学人工智能研究院常务副院长、计算机科学与技术系教授孙茂松，北大智能学院院长朱松纯，中国语言文学系主任杜晓勤，智能学院教授袁晓如，以及王军分别以“北京大学汉画研究所的计算艺术史探索”“更上层楼：

人工智能助力数字人文”“人工智能助力北大新文科建设”“中国古典诗歌声律分析系统及其应用举隅”“数字人文中的可视化”“作为信息基础设施的数字人文”为题发表了主旨演讲。

本次展览由北京大学主办，北大数字人文研究中心和智能学院可视化与可视分析实验室承办，艺术学院、视觉与图像研究中心、人工智能研究院和信息管理系协办，字节跳动公益给予了大力支持。展览将于3月17日—4月10日在北京大学全球大学生创新创业中心向全校师生开放。

1.7 “古籍智能信息处理”系列研讨会



为了更好地总结古籍智能化领域的前沿研究成果、方便来自不同地域、从事不同领域研究的专家学者针对古籍智能化进行深入交流、展望古籍智能化领域未来的发展方向和学术研究前景，北京大学数字人文研究中心、北京大学人工智能研究院共同主办了“古籍智能”系列研讨会。

2022年3月12日下午，“古籍智能”系列研讨会第一讲“智能时代古典文献学的机遇与挑战”正式在腾讯会议上举办，并在哔哩哔哩网站上同步直播。中国古籍保护协会刘惠平会长、全国高等院校古籍整理研究工作委员会秘书处卢伟秘书长、北京大学数字人文研究中心主任王军教授、北京大学中文系杨海峥教授、清华大学中文系刘石教授、浙江大学中文系徐永明教授、复旦大学中国历史地理研究所张晓虹教授、四川大学中国文化全球传播大数据中心王兆鹏教授、华南理工大学电子与信息学院金连文教授、韩国庆星大学韩国汉字研究所许喆教授、上海图书馆上海科学技术情报研究所刘炜研究员、南京师范大学文学院李斌副教授、北京大学中文系李林芳助理教授，以及其他在线的高校师生、古籍收藏和整理的

相关工作者等共计八十余人参与了研讨会，网上同仁共计一千三百余人观看了直播。

本次研讨会共设立四大主题，分别是：1) 智能信息环境下古典文献领域应用的技术、工具和平台；2) 古典文献研究的新拓展和新方向；3) 古籍的整理与出版将会迎来哪些改变；4) 古典文献的教学将因之产生哪些新的要素，人才培养方面又将出现哪些变化。在多元碰撞中共同思考古典与现代、传统与新术、人文与科技间是什么样的关系，又是被何种纽带所牵连。主题内容涵盖古典文献数字化技术研发、工具更新、平台建设、应用普及、学术研究、人才培养等多个维度。

2. 跨学科合作、校企合作与国际合作

2.1 与哈佛大学的全面国际合作



北京大学数字人文研究中心和哈佛大学费正清中国研究中心(FCCS)经过近一年的磋商，在12月底正式签署全面合作协议，内容包括：

- 1) 面向全球招收和培养数字人文领域的博士后研究人员，博士后在站期间一年在北大一年在哈佛。
- 2) 共同举办数字人文国际暑期学校及数字人文国际会议，第一届暑期课将于2021年7月在北京大学校园举办；第二届将于2022年在哈佛举办，随后在两校之间轮替。
- 3) 双方联合开展数字人文领域的数据集建设和课题研究。目前，双方联合申请的项目已经得到中国国家自然科学基金国际重点合作项目的支持。

- 4) 博士生交换学习：北京大学数字人文中心每年可以向哈佛费正清中国研究中心推荐两名博士生访问学习，期限一年。

上述全面合作协议的达成，为我校数字人文的发展奠定了国际化发展的基础，也标志着北京大学数字人文中心已经成为国际数字人文领域的重要力量。

2.2 字节跳动公益捐赠古籍数字化研究



3月17日，字节跳动向北大教育基金会提供捐赠，全面支持“北京大学-字节跳动数字人文开放实验室”的工作，研发古籍数字化平台，利用智能技术加速中华古籍资源的数字化建设，向全社会提供公益化服务。

以“北京大学-字节跳动数字人文开放实验室”为研发基地构建的古籍数字化平台。该平台将通过OCR（光学字符识别）、句读、实体识别、知识图谱构建等方面等多种技术，实现古籍的智能化整理，让古籍能够以文本的形态加以检索、关联阅读和深度挖掘与利用。该平台是一站式古籍智能化整理平台，将对社会大众免费开放。古籍收藏者、研究者、相关专业的师生和广大古籍爱好者可以利用该平台，展开古籍文字识别、句读、校对等全系列工作，提高古籍数字化的效率和质量。

“北京大学-字节跳动数字人文开放实验室”将调动国内外的相关力量，在三年内完成一万种精选古籍的智能化整理工作，基本覆盖儒家、道家和佛学的核心典籍目录。

2.3 校内跨学科的交流与研讨

数字人文中心与历史学系、中文系、哲学系、考古文博学院、外国语学院、艺术学院、信息科学技术学院、软件与微电子学院、城市与环境学院、地球物理与空间技术学院、信息管理系、图书馆的教师组织了多场跨学科研讨会，在我校形成了一个跨学科的松散学术联合体，成为我校跨学科交叉的一个学术桥梁。2020年期间举办的主要学术研讨活动有：

- 2020年1月16日上午，北京大学数字人文研究中心组织召开了主题为“人工智能时代人文学科的挑战与机遇”的数字人文建设与发展专题研讨会。来自中文系、历史学系、哲学系、考古文博学院、外国语学院、艺术学院、信息科学与技术学院、地球与空间科学学院、软件与微电子学院、信息管理系和图书馆等校内单位的学者就北京大学数字人文建设发展相关的需求、规划和跨界合作等进行深入探讨。学校社会科学部、研究生院、教务部、科学研究院和学科建设办公室相关负责人出席会议。通过这次研讨会，筹建中的数字人文研究中心收集了全校数字人文的建设需求，明确了建设目标，建立了与各院系的广泛联系渠道，为我校数字人文的发展指明了大致方向。



- 2020年5月下旬在数字人文工作坊连续六场在线授课之后，数字人文中心组织了线上研讨会，广泛听取了来自艺术学院、体教部、城市与环境学院、外国语学院、中文系、历史学系等院系的师生们的意见，收集了这些老师和同学在各自学科应用数字方法的需求和困难，对工作坊的教学成果进行了总结。

- 2020 年 9 月初，数字人文中心牵头组织了面向人文社科学生的人工智能课程建设研讨，联络信息技术学院李文新老师带领的人工智能授课团队，以及历史学系何晋老师、中文系邵燕君老师共同商讨联合授课事宜，由此促成了人工智能课程与历史系、中文系的联合授课。课程采用项目驱动的方式，三方调动了十余名研究生助教参与课程实践，经过一个学期的课堂讲授，形成了 4 个富有创新精神的研究课题，计划在本学期末汇报课程成果。
- 2020 年 10 月下旬，为做好国家 2021-2030 年古籍规划编制工作，古联公司受全国古籍整理规划小组办公室的委托，专程赴北京大学，征集建议项目及古籍数字化相关意见。数字人文中心组织了本次古籍数字化项目调研会。来自历史系、中文系、《儒藏》编纂与研究中心、外国语学院、图书馆等单位的教师近 20 人参加。老师们从古籍数字化的标准制定、古籍整理平台、古籍相关研究工具等方面提出了自己的需求与建议。
- 2020 年 11 月 13 日，数字人文中心与城市与环境学院王长松老师带领的历史地理团队交流了双方在数字人文领域的研究成果。王长松老师组织历史地理方向的博士生和硕士生参加了上学期的数字人文工作坊，并应用工作坊介绍的数字人文工具进行了一系列的研究。在交流会上，历史地理方向的同学们介绍了应用 GIS 空间分析和词汇标注工具所展开的若干研究项目，包括：“唐代河北道城市体系发展及其影响因素研究”、“想象的城市：南京诗词景观的塑造与网络结构”、历史时期黄河流域城市空间格局演变与影响因素”等；数字人文实验室的同学们介绍了“《朱子年谱》时空叙事可视化方法研究”、“《宋元学案》知识图谱系统平台”、“数字人文视角下唐代诗人研究”等研究成果。
- 2021 年 1 月 14 日上午，由北京大学数字人文研究中心主办的“智慧人文的探索与实践”研讨会暨北京大学数字人文研究中心成立仪式在北京大学英杰交流中心月光厅顺利召开。北京大学数字人文研究中心学术委员会委员历史学系邓小南教授、信息科学技术学院李文新教授、中国语言文学系刘玉才教授、外国语学院赵白生教授共同为北京大学数字人文研究中心揭牌。“智慧人文的探索与实践”研讨会上半场以“人工智能在人文领域的教学、研究与实践”为主题。研讨会下半场以“数字人文视角下的人文研究”为主题。来自智能科学系罗定生副教授、外国语学院苏祺副教授、历史学系何晋教授团队、信息管理系王军教授团队、软件与微电子学院俞敬松副教授、中文系邵燕君副教授团队、城市与环境学院王长松副研究员、体育教研部赫忠慧教

授、《儒藏》中心杨浩助理教授等分别分享了各自领域与数字人文相关的前沿研究。在讨论环节，参会的老师们进行了热烈的讨论。最后，北京大学历史学系朱青生教授通过线上表达他对北京大学大力开展跨学科融合实践的期望。



- 2022年1月13日下午，由北京大学数字人文研究中心主办，北京大学文化产业研究院和北京大学文化传承与创新研究院联合承办的跨学科学术沙龙在北京大学燕南园51号院举行。此次学术沙龙活动以“文化遗产与文化计算”为主题，来自信息管理系、信息科学技术学院、文化产业研究院、社会科学部、法学院、哲学系、考古文博学院、王选计算机研究所、地球与空间科学学院、中古史中心、外国语学院、《儒藏》编纂与研究中心等北京大学不同学科领域的20多位学者参加。本次沙龙活动由史睿主持。由信息管理系教授、数字人文研究中心主任王军、文化产业研究院院长、文化传承与创新研究院常务副院长向勇、社会科学部基地与机构办公室主任郭琳致欢迎辞。沙龙的主体环节是专题分享。首先由北京大学数字人文中心苏祺、杨浩介绍了目前数字人文中心的主要工作，即围绕古籍所作的一些利用与开发，通过AI技术，从普及最开始的原始形态，最后组织成一个知识图谱。接着，信息科学技术学院教授袁晓如在专题分享中介绍了实现数据可视化的技术手段，包含可视化技术、交互技术、数据结构、数据库检索等多种技术。王选计算机研究所教授周秉锋介绍了虚拟现实中的绘制技术的成果应用。最后，中古史中心副研究员史睿强调不同领域的研究文献、研究文物、研究图像的成果需要汇总，有利于实现文化遗产知识的有序性和传播的有效性。专题分享后，

各位学者围绕“数字人文、数字可视化、知识图谱”等进行了积极互动。大家认为，文化遗产与文化计算作为数字人文的重要研究方向之一，是新文科背景下科技与人文交融的集中表现，展现了融合传统文学、历史学和计算机科学、图书情报等跨学科研究的独特魅力。



3. 课题研究

■ “国家古籍数字化工程”专项经费资助项目

由全国古籍整理出版规划领导小组规划的 2021 年，全国范围内共收到 300 多项申请。经过数轮严格评审最后通过 27 项，通过率仅 9%。其中，北大数字人文中心主持、协作和参与的古籍数字化工程项目共有 5 项，占到所有资助项目的 18.5%。具体列举如下：

- 1) 面向典籍文本的知识图谱自动生成平台，北大数字人文中心独立申请。
- 2) 历代典籍目录数据集成与分析系统，北大数字人文中心独立申请。
- 3) 国家珍贵古籍名录知识库建设，国家图书馆和北大数字人文中心联合申请。
- 4) 《永乐大典》高清影像数据库，国家图书馆出版社独立申请，全权委托北大数字人文中心主持设计和开发。
- 5) 《儒藏》精华编数字化平台，北大出版社申请，数字人文中心杨浩老师参与撰写。

■ 国家自然科学基金国际重点合作项目“中国儒家学术史知识图谱构建研究”

北京大学数字人文研究中心与哈佛大学费正清中国研究中心联合申请的国家自然科学基金重点国际合作项目“中国儒家学术史知识图谱构建研究”获得资助，项目的执行期限是 2021 年 1 月至 2025 年 12 月。课题的目标是应用深度学习的方法从中国历代儒家典籍中提取命名实体和语义关系，在古代地理和历史知识本体的指导下自动构建儒家学术史知识图谱，并开发一系列可视化、社会网络、统计分析等工具来呈现 2500 年来儒家学术的全景式流行和学说思想的概念演化，实现一个集教学、研究与文化传播为一体的国际化数字人文研究平台。

■ 中宣部出版局委托项目“古籍数字化关键技术创新与应用”

2020 年 6 月，数字人文中心承担中宣部出版局全国古籍规划管理办公室的委托项目“古籍数字化关键技术创新与应用研究”。我们组织了北大、北师大、复旦三所高校的八名老师组成联合研究团队，完成课题的调研任务，并撰写了调研报告 5 万余字。在此基础上，组织全国十三家古籍数字化相关机构成功召开了“智能环境下古籍资源开发与利用”的研讨会。

■ 中美合作“中国历史人物资料库”(CBDB)在线查询系统

北京大学数字人文研究中心为哈佛大学费正清中心主持建设的“中国历史人物资料库”设计并研发在线查询系统的升级版，完善现有查询功能存在的不足，重新构建一个对用户友好的 CBDB 在线查询系统。新系统提供了更加清晰的交互逻辑，并支持资料的导入与导出，具有界面统一、支持多语言界面、使用便捷、资料库及时更新的特点，符合人文学者在研究中对 CBDB 数据查询的需求。新的 Web 查询系统即将部署在哈佛费正清中国研究中心的“中国历史人物资料库”的主服务器和北京大学图书馆的服务器上，面向全球用户开放，服务更多的中国古代史研究人员。这个项目的核心开发人员陈诺同学赢得了 2020 年度“北京市本科优秀毕业论文”的荣誉。

4. 论文发表

北大数字人文中心在与数字人文相关的各级国际会议、SCI 和 SSCI 源刊、

CSSCI 国内期刊上共发表学术论文近 30 篇。内容涉及：古文自然语言处理技术、本体工程和知识图谱在古籍领域的应用、数字人文的发展趋势、数据驱动的文化分析、人文学者的信息需求研究、宋元明清学术谱系的可视化分析、数字人文软件系统和平台构建、人工智能领域的文献计量等。中心构建的数字人文软件系统在全国数字人文会议上获得最佳项目奖和最佳创意奖。

在短短两年内取得研究成果，表明数字人文作为一个交叉学科阵地，有着巨大的学术发展潜力。这些成果的发表也使得北大数字人文中心成为有国际学术影响力的研究阵地。发表情况列举如下：

■ ACL 2021 (国际计算语言学年会 , EI)

Xuemei Tang, Jun Wang, Qi Su. That Slepen Al the Nyght with Open Ye! Cross-era Sequence Segmentation with Switch-memory[C]. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics(ACL).

■ CCL 2021 (中国计算语言学大会)

唐雪梅, 苏祺, 王军, 陈雨航, 杨浩. 2021 .基于预训练语言模型的繁体古文句读研究[C].第 20 届中国计算语言学大会论文集:678-688.

■ CIKM 2021 (ACM 国际信息与知识管理会议 , EI)

Siyu Duan, Wei Li, Jing Cai, Yancheng He, Yunfang Wu. Query-Variant Advertisement Text Generation with Association Knowledge. Proceedings of the 30th ACM International Conference on Information & Knowledge Management (EI)ISBN-13:9781450384469; DOI:10.1145/3459637.3482290;

■ JCDL 2021 (数字图书馆联合会议 , EI)

LI W, WANG F, WANG J. Exploring the Classification of Traditional Chinese Bibliographies through Interactive Visualization[C]//2021 ACM/IEEE Joint Conference on Digital Libraries (JCDL). . DOI:10.1109/JCDL52503.2021.00071. (EI, NSFC)

■ iConference 2021 (国际信息学院联盟年会 , EI)

Chengxi Yan, Tzu-Yi Ho, and Jun Wang, A Comparative Studies of Automatic Query Formulation in Full-Text Database Search of Chinese Digital Humanities,

iConference 2021: Diversity, Divergence, Dialogue pp 457-468
Tzu-Yi Ho, Chengxi Yan, and Jun Wang. Exploring Query Formulation in Digital Humanities: A Preliminary Study on Visual Query Expansion in Chinese Historical Databases. iConference 2021: Diversity, Divergence, Dialogue pp 457-468

■ CDH 2021 (中国数字人文大会)

王林旭, 位通, 王军. 基于 OWL-Time 的时间本体在古籍中的应用. 第三届中国数字人文大会,南京,2021. (优秀论文二等奖, NSFC)
李佳纯, 杨浩, 苏祺, 王军. 论语在儒家典籍中的文本复用计量研究. 第三届中国数字人文大会,南京,2021. (优秀论文二等奖)
李山雨, 季佳雯, 邓泽琨, 苏祺. 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 . 第三届中国数字人文大会,2021.

■ International Conference on Big Data 2020(大数据国际会议)

Yan, C., & Wang, J. (2020). Exploiting Hybrid Subword Information for Chinese Historical Named Entity Recognition. 2020 IEEE International Conference on Big Data (Big Data), 4795-4801.

■ DADH 2020(台湾第 11 届数字人文与数字典藏国际会议)

王林旭, 王軍. 基於宋元明清儒學史的學術譜系研究. 第十一屆數位典藏與數位人文國際研討會(DADH 2020, International Conference of Digital Archives and Digital Humanities), 台北, 2020.

王凤翔 王军. 基於知識圖譜的《論語》視覺化系統設計與構建. 第十一屆數位典藏與數位人文國際研討會(DADH 2020, International Conference of Digital Archives and Digital Humanities), 台北, 2020

陳諾, 卞恩華, 王祥安, 盧建安, 王宏甦, 王軍. 面向使用者需求的 CBDB 線上查詢系統設計與實現, 第十一屆數位典藏與數位人文國際研討會(DADH 2020, International Conference of Digital Archives and Digital Humanities), 台北, 2020.

■ DH 2020 (ADHO 数字人文国际会议)

Jun Wang, Liyuan Zhang, Haici Yang. Can Digital Humanities Help in Finding Research Questions? A Comparative Analysis of the Attitudes Towards Neo-Confucianism Study of the Scholars Today and 300 Years Ago. Digital Humanities Conference, 2020.

Jun Wang, Xiaoyu Li, Enhua Bian, Linxu Wang, Shuran Liu, Nuo Chen. A Visualization-Assisted Reading System for a Neo-Confucian Canon. Digital Humanities Conference, 2020.

■ JCDL 2020 (数字图书馆联合会议)

- Shuran Liu, Jun Wang. How to Organize Digital Tools to Help Scholars in Digital Humanities Research?. Joint Conference on Digital Libraries, 2020.

■ 期刊论文

Tong Wei, Yuchen Sang, Shufen Chen, Letian Wang, Rui Wang, Qi Huang, Jun Wang*. WebGIS Approach of Entity-oriented Search to Visualize Historical and Cultural Events. Digital Scholarship in the Humanities, 2022. DOI: <https://doi.org/10.1093/lhc/fqac002> (SSCI, NSFC)

Fengxiang Wang, Tong Wei*, Jun Wang. Ontology-Driven Cultural Heritage Conservation: A Case of the Analects of Confucius. Applied Science, 2022. DOI:<https://doi.org/10.3390/app12010287> (SCI, NSFC)

Wei T, Roche C, Papadopoulou M, Jia Y. Using ISO and Semantic Web standard for building a multilingual terminology e-Dictionary: A use case of Chinese ceramic vases. Journal of Information Science. June 2021.
doi:10.1177/01655515211022185 (SCI)

Wei T, Roche C, Papadopoulou M, Jia Y. A Term-and-Characteristic Guided Methodology for Building Ontology of Cultural Heritage. ACM Journal on Computing and Cultural Heritage. 2022 (拟接收, SCI, NSFC)

Wei T, Roche C, Papadopoulou M, Jia Y. The TAO CI Ontology of Vases of the Ming and Qing Dynasties. Applied Ontology. 2022.(拟接收, SCI, NSFC)

Tong Wei, Christophe Roche, Maria Papadopoulou, and Yangli Jia. Semantic Web and ISO Standards to Create Multilingual Terminology e-Dictionary for Cultural Heritage: A Mini Review. International Journal of Sensor Networks and Data

Communications. 2021, Volume 10:S5. (已接收)

C. Yan, Q. Su and J. Wang, "MoGCN: Mixture of Gated Convolutional Neural Network for Named Entity Recognition of Chinese Historical Texts," in IEEE Access, vol. 8, pp. 181629-181639, 2020, doi: 10.1109/ACCESS.2020.3026535.

位通, 桑宇辰, 史睿, 王军*. 基于知识重构的年谱古籍的时空可视化呈现——以《朱熹年谱长编》为例.[J/OL]中国图书馆学报, 2022.DOI:10.13530/j.cnki.jlis.(NSFC)

李文琦, 王凤翔, 孙显斌, 黄芷欣, 李芃蓓.历代史志目录的数据集成与可视化.[J]中国图书馆学报 (已接收, CSSCI, NSFC)

王林旭, 杨浩, 王军.基于学案体文献的儒家学术史可视化分析与呈现[J]. 中国图书馆学报, 2022 (已接收, NSFC)

张力元, 毕梦曦, 王军. 当代宋明理学研究热点与演变趋势分析, 数字人文, 2022 年第 2 期 (已接收)

苏祺,胡韧奋,诸雨辰,严承希,王军.古籍数字化关键技术评述[J].数字人文研究,2021,1(03):83-88.

张力元,王军. 基于时态社会网络的两宋学术和政治体系比较分析.情报工程, 2020, 6(01):34-49.

张力元,王军. 古籍数据库分面分类体系设计[J/OL].图书馆建设[2020-11-17].

王军. 从人文计算到可视化——数字人文发展脉络梳理, 文艺理论与批评.2020 年第 2 期

王军,张力元. 国际数字人文进展研究.《数字人文》创刊号.2020 年第 1 期

撰稿:王军、杨浩

封面设计:王瑞



扫码关注公众号
数字人文开放实验室