201911 北京数字人文兴趣小组研讨会实务手册

201911 DHSIG - Beijing Seminar

Practice Manual

**DHSIG – Beijing 研讨会管理委员会**

**王宏甦 徐力恒 徐阳 张力元 汪伟 高树伟**

本次会议主题

**人文学者与非人文学者如何在数字人文研究中合作**

研讨会简介与目标

近几年数字人文在中国发展迅速，讲座、工作坊、论坛等成为介绍、讨论数字人文的重要形式。随着数字人文基本知识、方法论和工具的普及，如何利用这些知识、方法论、工具切实地解决人文领域的学术问题，如何推动数字人文知识、方法论、工具本身的发展，逐渐成为大家越来越迫切的需求。我们希望以例会的形式在北京大学人文社会科学研究院举办的“北京数字人文兴趣小组研讨会”（下文简称 DHSIG 研讨会）正是对此需求的积极回应。

DHSIG 研讨会的目标为：定期对数字人文研究议题、热点、技术、思潮组织圆桌讨论。讨论不设主讲者，只设置引导讨论的主持人/主持小组。视所有参与者为平等的研讨者，会前向参与者分发相关材料，以便讨论者在会前了解此主题的基本背景。为促进学生积极发言，每个环节的议题我们先邀请学生发言，之后所有与会者再加入讨论。以期通过研讨会成员的讨论，对讨论对象形成共同结论。会后共同署名向公众发表这一结论（这一结论内部可能由多种互补，甚至对立的论点组成）。我们希望这一结论或在纷繁的讨论中，为数字人文研究者对某一议题提供一系列可操作的方案；或将当前的某些探讨、探索面向未来推进。

为了实现这一目标，我们希望：

1. 所有参会者，无论是老师、同学，在不偏离会议主题、不超时的前提下，充分、平等地发表见解。我们不欢迎不做会前准备，及会上不表达观点的“听众”；
2. 所有参会者会前阅读“实务手册”及“实务手册”中涉及的相关材料，以便了解会议的议题、相关文献，以及由委员会起草供大家讨论的“公开报告参考关键词”；
3. 对“实务手册”有任何建议或者补充，DHSIG 研讨会成员可随时在微信群或 DHSIG - Beijing 论坛中进行实时讨论；
4. 为帮助研讨会中学生充分表达意见，在环节的设置上，我们先会邀请博士生、硕士生、本科生对该环节的议题进行第一轮讨论，再由老师、博士后、图书馆专家、从业专家等加入，进行第二轮自由讨论。
5. 会议结束后，我们会整理大家的观点，以全体参会人员署名的方式向公众发表：1）会议的 “结论” 。2）本次会议的 “实务手册” 。3）本次会议录音中最高频的非 stop words、非隐私信息的高频关键词。4）本次会议涉及的“参考文献” 及会上、会下提及的其他重要可公开资源。

时间与地点

11/6 10:00-11:30am, 北京大学人文社会科学研究院 111 会议室

背景知识：数字人文研究方法

从不同角度观察，数字人文常见的研究方法非常多样。包括中国历代人物传记资料库在网站上汇总的方法论：群体传记学、地理信息系统、社会网络关系分析 <https://projects.iq.harvard.edu/chinesecbdb/%E6%96%B9%E6%B3%95%E8%AB%96>。包括黄一农教授提出的“E考据” <https://baike.baidu.com/item/E%E8%80%83%E6%8D%AE>，方法实践见参考文献中 New Frontiers of Electronic Textual Research in the Humanities: Investigating Classical Allusions in Chinese Poetry through Digital Methods。

中国历代人物传记资料库项目清华小组基于近年 DADH 和 Alliance of Digital Humanities Organizations 的 DH 会议，对数字人文相关技术做了综述。<https://github.com/cbdb-project/overview-of-DH-methods-for-Asian-studies/blob/master/CBDB%E6%B8%85%E5%8D%8E%E5%B0%8F%E7%BB%84%E6%95%B0%E5%AD%97%E4%BA%BA%E6%96%87%E7%A0%94%E7%A9%B6%E6%96%B9%E6%B3%95%E7%BB%BC%E8%BF%B0_final.docx> 综述中涉及到的技术有：主题模型（Topic Model）、序列联配 (Sequence Alignment)、中文分词 (Chinese Word Segmentation)、主成分分析（Principal Component Analysis）、聚类算法（Cluster）、支持向量机（Support Vector Machine）、词夹子方法（Word-Clip）、词嵌入方法（Word Embedding）、社会网络分析 (Social Network Analysis)、地理信息系统 (Geographic Information System)、数据库提升 (Database Upgrade) 。

会议环节

10:00-10:15 人文学者希望用数字人文的方法解决什么问题？

10:15-10:30 数字人文研究中，人文学者和非人文学者在交流中需要注意什么？

10:30-10:45 数字人文研究中，人文学者和非人文学者应该接受哪些训练？

10:45-11:00 如何评估数字人文研究的价值？

11:00-11:15 自由讨论

11:15-11:30 形成哪些结论？

公开报告参考关键词

**Topic**: 人文学者希望用数字人文的方法解决什么问题

**Keywords**: 节约时间、纸质载体的限制、复杂研究方法的自动化

**Topic**: 数字人文研究中，人文学者和非人文学者在交流中需要注意什么

**Keywords**: 研究目标和方法、专业知识、专有名词、充分讨论、不相互干涉的界限、署名

**Topic**: 数字人文研究中，人文学者和非人文学者应该受到哪些训练

**人文学者训练 Keywords**: 正则表达式，地理信息系统分析，社会网络关系分析，关系型数据库，文本编码，Excel 基本公式和操作，python 自然语言处理库，基于数据的思想

**非人文学者训练 Keywords**: 普适性训练，专题训练

**Topic**: 如何评估数字人文研究的价值

**Keywords**: 人文领域价值、非人文领域价值、工程性、开放性

参考文献

1. "Chen Song: Why Humanists Should Fall in Love with “Big Data,” and How?(为何人文科学学者应该爱上大数据？How？).
2. 徐力恒: 一个新领域的出现：中国的数字人文研究.
3. 陈佩辉.人文数据库建设中人文学者何为——以《全宋文》墓志铭亲属信息提取为例[J].图书馆论坛,2019,39(05):17-23.”
4. 王涛 什么不是数字人文, 澳门理工学报201904.
5. 中国历代人物传记资料库方法论 <https://projects.iq.harvard.edu/chinesecbdb/%E6%96%B9%E6%B3%95%E8%AB%96>
6. “E考据” 百度百科. <https://baike.baidu.com/item/E%E8%80%83%E6%8D%AE>.
7. Huang, Yi-Long, and Bingyu Zheng. "New Frontiers of Electronic Textual Research in the Humanities: Investigating Classical Allusions in Chinese Poetry through Digital Methods." Journal of Chinese Literature and Culture 5, no. 2 (2018): 411-37.
8. CBDB清华小组数字人文研究方法综述 <https://github.com/cbdb-project/overview-of-DH-methods-for-Asian-studies>

参考文献下载地址：

<https://github.com/DHSIG/dhsig_bj_files/tree/master/Seminar%20references/201911/references>

微信群与线上论坛

研讨会微信群二维码：



或者添加王宏甦(微信 ID: sudomail)为微信好友，王宏甦会把您添加到 DHSIG – Beijing 研讨会微信群中。

研讨会论坛：

[https://dhsig.oopus.info](https://dhsig.oopus.info/)

个人信息提交

<https://www.wjx.cn/jq/47876808.aspx>

致谢

本研讨会思路源于北京大学中古史中心史睿老师提出的“数字人文倡议书”。没有这一原始动机，便不会有 DHSIG 研讨会。向史睿老师表示诚挚的感谢！

感谢北京大学人文社会科学研究院对本研讨会的全力支持！