

专题：面向人文领域的 大数据技术和方法

Big Data Technology and Method in Digital Humanities

客座编辑



袁晓如 (1975-)，男，博士，北京大学智能学院研究员、博士生导师，机器感知与智能教育部重点实验室副主任，大数据分析与应用技术国家工程实验室常务副主任，中国计算机学会杰出会员，主要研究方向为可视化与可视分析。现任《大数据》编委，中国图象图形学学会理事、可视化与可视分析专业委员会主任。

导读

近年来,大数据技术在全球取得了巨大的发展,深刻改变人类社会生活和运行的各方面。同时,传统学科也开始与数据技术和方法进一步紧密结合。在人文方面,数字人文的概念得到了更多的关注,大数据管理、分析、可视化等技术被大量应用于人文研究和探索,人文领域跨学科交叉研究越来越受到重视。本刊“面向人文领域的大数据技术和方法”专题,对涉及的方法、技术、现状和挑战等方面进行深入讨论。

南京大学陈静系统分析了人文大数据的特点,梳理了人文大数据与科学、工程以及社会科学数据的差异,分析了相比科学计算、工程计算以及社会计算,人文计算的特殊性及其价值,结合数字人文研究案例讨论了人文大数据研究应用的价值和可能存在的挑战。

数字人文工作的一个重要方向是古籍分析与处理。由于古籍语法结构和语言特点特殊,大规模古籍资源的标注是数字人文工作的难点和痛点。华东师范大学于亚秀等人探索了基于MARKUS的文本标注模型,对进一步建立普适性的古籍标注标准流程具有重要意义。天津大学韩立帆等人基于预训练模型对历史古籍中的实体和关系等进行了知识抽取,并进一步构建了一个面向历史古籍文本标注任务的众包系统,对于有效获取历史古籍文本中的丰富信息

具有很大意义。南京师范大学与南京农业大学的郑童哲恒等人以《史记·列传》为对象,进行古汉语自动分词及词性标注、人工校对以及实体信息人工标注,形成多层次、高质量的数字人文知识库,实现对古籍词汇、人物、地点等要素的定量分析与可视化检索,为古籍数字人文知识库的构建提供新的思路与框架。

本专题部分文章关注与数字人文相关的一些技术。四川轻化工大学与西南科技大学的张伶俐等人对文本情感可视分析技术做了系统的梳理,从主题词、关联、演变、时空分布等方面阐述文本情感可视分析方法,对研究如何利用情感分析技术,特别是文本情感可视分析有所帮助。北京大学的罗煜楚等人总结了当前可视化在数字人文中应用的现状,多角度对数字人文项目进行分析,揭示多学科交叉对促进数字人文与可视化结合的巨大潜力,同时分享了面向数字人文需求的可视化的多学科交叉合作新范式的实践。

受篇幅限制,本专题难以涵盖数字人文领域的全部内容。这一研究方向越来越受到国内外各个研究机构的重视,希望本专题可以使来自不同领域的读者更全面地认识与数字人文相关的技术和应用发展的问题,促进跨学科的合作,共同推动数字人文发展。