图书馆论坛
Library Tribune
ISSN 1002-1167,CN 44-1306/G2

《图书馆论坛》网络首发论文

题目：作为数字人文基础设施的图书馆：从不可或缺到无可替代
作者：刘炜
网络首发日期：2019-12-13
引用格式：刘炜．作为数字人文基础设施的图书馆：从不可或缺到无可替代[J/OL]．图书馆论坛. http://kns.cnki.net/kcms/detail/44.1306.G2.20191213.0827.008.html

作为数字人文基础设施的图书馆：从不可或缺到无可替代

刘炜

根据 CNKI 数据，国内迄今发表的近 700 篇以“数字人文”为主题的论文中，来自图书情报档案领域的文章约超过 60%。对比国外，Web of Science （Core Collection）收录了 1,590 篇以“digital humanities”为 topic 的论文，Inforamtion Science Library Science 领域的文章约 300 篇，占比不到 20%。这两组数据显示国内外数字人文研究学科来源的巨大差异。这说明了什么呢？虽然我们并不认为国外的比例就是数字人文知识版图的“完美”配方，但我们的比例一定是不合理的。人文学科的数字疆域，第一批居民主要来自图书情报领域，怎么说都不能让人服气。这其中固然有国内图书情报学者更喜欢追新的原因，也是国内人文领域的学者尚未觉醒、尚未充分准备好的结果。就像当初旧金山发现了金矿，涌入的首批淘金者并没有赚到钱，而各类服务业却异军突起。图书馆行业作为历史文献的主要保留地，由于数字图书馆带来先知先觉，理所当然地成为数字人文最早的基础设施建设者。

传统的文献考据和现代的文献计量学都为数字人文作为一个整体的跨学科研究领域提供了方法论借鉴，书目控制带来的规范控制借助于语义技术，天然地为知识的形式化组织（采用本体技术）和知识服务提供了可信的编码基础，也为机器学习和人工智能的发展提供了宝贵的标注语料库。如果说不了解目录之学就无法窥知传统学术门径的话，不懂得以文献计量为代表的统计分析方法就无法真正从事数字人文研究。当然，当今数字人文的方法体系已经得到了极大拓展，统计分析的对象从文献深入到了语词文本、社会关系、时空关系乃至经过模型化之后的各类关系。但无论多么复杂，数据永远是基础，拥有大量数据的图书馆永远是人文研究的可靠伙伴。

图书馆要提供基于知识的服务还需要在数字图书馆的基础上不断提升水平，包括提升资源加工的语义化水平和提供分析统计及可视化工具。上海图书馆在国内属于对数字人文的先知先觉者之一，借助于 20 多年前开始的持续不断的数字化，大量的传统文献和特色文献已经搬运到了数字世界，一旦数字人文的研究方法和相关技术得以成熟，很自然地占据了有利的跑道。

本专题的 4 篇文章虽然反映不了上海图书馆在数字人文领域积极开拓的全貌，但包含了一些新的思考。图书馆这类人类记忆机构在数字人文的发展过程中，固然由于其资源收藏而不可或缺，但真正使其无可替代的，并不是这些馆藏资源，而是其服务能力。在当今以 ABCD（人工智能、区块链、云计算和大数据）为特征的数字时代，“知识作

为一种服务（KaaS）”才是图书馆的立身之本。本专辑反映了数字人文平台建设的两大趋势：边服务边建设的开放众包思想，和从数字图书馆到数据图书馆的必要升级。这两者是“后数字图书馆时代”我们在面向数据驱动型或数据密集型研究进行转型时必须首先实现和超越的。

贺晨芝和张磊的《图书馆数字人文众包项目实践》重点介绍了数字人文领域的众包应用现状，以及上海图书馆自 2016 年以来的实践经验。上海图书馆开发了两个独立的众包应用，即抄录平台和验证码应用，都可以以 SaaS 方式开放给同行使用。

刘倩倩和夏翠娟的《家谱知识服务平台众包模式的设计与实现》针对上海图书馆的家谱特藏，在原来提供基本查询和关联功能的数字人文平台基础上，开发了上传家谱、在线识谱、在线修谱等功能，尝试引入众包模式不断优化系统，并与用户社区积极互动密切合作，使用户不仅作为数据的消费者，同时也作为贡献者。

朱武信和夏翠娟的《命名实体识别在数字人文中的应用—基于 ETL 的实现》介绍了一种借助于专门词典、批量自动进行名称实体识别的方法（即 ETL 方法），该方法在上海图书馆的数字人文平台建设中已普遍采用，取得了良好的效果。其原理是将文本中有意义的名称（例如人物、地点、时间、事件、专有概念等）利用程序进行自动析取，经过判断之后进行数据数据化转换（通常是加上 URI），并提供丰富的语义关系。

张喆昱和张磊的《记忆机构开放数据建设及数据化转型模式研究》触及了两个关键性主题：数据化和开放服务，试图将上海图书馆的实践进行一般化和通用化，分析了如何通过数据化更加贴近人文学者的需求，让系统更加人性化，然后通过开放服务引入外部资源，反过来促进系统的数据化。

上海图书馆希望通过自己的实践，为人文研究的赛百基础设施建设提供一个参考样本。发表这些做法，并不是说我们的做法有多先进，而只是一种不揣浅陋的抛砖引玉。我们深知，国内的数字人文目前还处于起步阶段，争论大于共识、口水多于实践，但只要大家参与，未来就前景可期。数字人文迄今为止形成的最大共识，就是大家都同意它是一个人人都受欢迎从而能各得其所的“大帐篷”。愿这个大帐篷使我们各门人文学科都得到繁荣兴旺！

作者简介：刘炜，博士，上海图书馆副馆长。

（责任编辑：刘洪）