Humanistic inquiry with large corpora of digitized text and metadata: Towards new epistemologies?

Workshop Proposal (short version)

Workshop, 130th Annual Conference of the Modern Language Association (MLA), 8–11 January 2015, Canada. (proposal submitted, April 1, 2014).

Sayan Bhattacharyya and Jeremy York

The availability of large corpora of digitized text from the world’s research libraries has the potential to inform humanistic inquiry in epistemologically novel ways. Not only scholars who specialize in digital humanities but also those who pursue traditional scholarship can potentially benefit from using the resources and tools that are now becoming available in this domain. Our workshop will be grounded in a particularly interesting and voluminous research corpus, the HathiTrust corpus, which consists of more than ten million (and growing) books (from some of the world’s foremost research libraries, which have joined the non-profit HathiTrust consortium), along with the bibliographic metadata associated with these books. This affords the ability to scale up inquiry in unprecedented ways and also allows new kinds of research questions to be asked. Even when copyright restrictions preclude the availability of full-text content to scholars, innovative methods of analysis using easy-to-use, but sophisticated, computational tools, such as the ones that the HathiTrust’s research-oriented affiliate, the HathiTrust Research Center, has been engaged in developing, can open up new avenues for humanistic inquiry.

While using the HathiTrust and the HathiTrust Research Center’s activities, collection and tools as an exemplar to keep the discussion grounded by providing an implemented, deployed and demonstrable working system, our workshop, however, aims to take the discussion to a more theoretical level than mere how-to instructions. The principal objective of the workshop will be to discuss the new questions and challenges that the arrival of these novel methods are beginning to generate in connection with the epistemology of humanistic inquiry. The two presenters, who combine a background in the humanities with familiarity with computational tools, will speak, for about 10 minutes each, about the collection, the metadata and the tools associated with our ground instance, and then provide a 15-minute long illustrated demonstration (live, with screenshots as backup) to familiarize the participants with innovative methods and practices, and to broach the epistemological issues that are at stake. The remaining 25 minutes will be devoted to an intensive discussion involving all participants/attendees about the new thinking concerning such epistemological matters as the relationship between close and distant reading, the hermeneutics of digitized text, the place for collaborative production of knowledge, and the increased blurring of roles between scholars and librarians, and between curators, annotators/editors and interpreters of text, that these novel developments are likely to inaugurate. The target audience for the workshop is not only the specialist digital humanities scholar, but also, and more especially, the curious-minded and adventurous humanists at large. Last but not the least, our workshop is affiliated with the presidential theme of the MLA 2015 conference, “Negotiating Sites of Memory.” The collections of digitized text are sites of mankind’s cultural memory, and to negotiate them in a way that realizes their full potential for humanistic scholarship demands new thinking in the terrain of epistemology.