Glenn Roe

Mellon Fellow in Digital Humanities

University of Oxford

http://users.ox.ac.uk/~fmml2030/

ABSTRACT

The challenge of 'Big Data' in the Humanities has led in recent years to a host of innovative technological and algorithmic approaches to the growing digital human record. These techniques—from data mining to distant reading—can offer students and scholars new perspectives on the exploration and visualisation of increasingly intractable data sets in the human and social sciences; perspectives that would have previously been unimaginable. The danger, however, in these kinds of 'macro-analyses', is that scholars find themselves increasingly disconnected from the raw materials of their research, engaging with massive collections of texts in ways that are neither intuitive nor transparent, and that provide few opportunities to apply traditional modes of 'close reading' to these new resources. In this talk, I will outline some of my previous work using data mining and machine learning techniques to explore large data sets drawn primarily from the French Enlightenment period. Building upon these past experiences, I will then present my current research project at Oxford, which uses sequence alignment algorithms to identify intertextual relationships between authors and texts in the 18th-century "Republic of Letters." By reintroducing the notion of (inter)textuality into algorithmic and data-driven methods of 'macro-anlalysis' we can perhaps bridge the gap between distant and close readings, by way of an intermediary mod e of scholarship I term 'directed' or 'scalable' reading.