The Quality in Quantity - Enhancing Text-based Research

1 Andreas ASCHENBRENNER, 2 Bernie. ACS, 3 Tobias BLANKE, 3 Mark HEDGES, 1 Felix LOHMEIER, 4 Wolfgang PEMPE & 4 Angus ROBERTS

1 Goettingen State and University Library, DE
2 National Centre for Supercomputing Applications, UIUC, US
3 King’s College London, UK
4 University of Sheffield, UK

Gregory Crane asked: "What Do You Do with a Million Books?" (http://www.dlib.org/dlib/march06/crane/03crane.html), illustrating a vision of how reference networks of books, annotations, and other sources are creating an additional layer of knowledge beyond the mere words. This vision is driven by the experiences in the arts and humanities for dealing with unstructured, context-dependent and often distorted information, yet it is also informed by statistics and other scientific methods. (http://www.acls.org/cyberinfrastructure/) Projects exploring the opportunities in text-based research often focus on either qualitative or quantitative mechanisms. For example, projects like TextGrid (http://www.textgrid.de) and TEXTvre (http://www.kcl.ac.uk/iss/cerch/projects/portfolio/textvre.html) establish virtual research environments for qualitative research in the humanities, allowing researchers to collate, edit and share texts. GATE and SEASR on the other hand provide frameworks for computational analysis of large amounts of data. There is relevant research for computational processing of textual resources (i.e. information retrieval, text mining), but less sustainable work has been done to consider respective infrastructural needs.

It is obvious that qualitative and quantitative approaches are complimentary. However, we need to better understand how existing tools and services can be brought together in one integrated research infrastructure
(1) through which distributed textual resources can be shared in secure environments;
(2) through which methods can be shared, co-developed, and discussed;
(3a) where researchers can manage, analyse and visualise large bodies of text and subsequently drill down into specific works (or "bags of words");
(3b) or vice versa edit a specific text by contextualising it in networks of related works and supplementary resources.

The opportunities in integrating existing methods go beyond research in the arts and humanities and also include existing approaches for text mining in biotechnology, machine assistance in annotation and contextualisation of documents in enterprise knowledge management systems, and others. The integration of quantitative and qualitative approaches will enhance the research opportunities for researchers in the digital humanities and stimulate new research by connecting the academic communities.