

ultimately be included or excluded from public connectivity which is enabled via broadband, WiFi and satellites. As transmission modes continue to partly inform the development of GPS location based applications and social media, it is time for a discussion that also includes notions of civic engagement.

Drones are now used for artistic, commercial and communication ends, as well as political purposes, seen in the works of Bridle. Like online social media and data driven journalism, drones with high definition cameras now also have the potential to change journalism practice. This has come about via the introduction of the BBC's Hexacopter, which produces different visual perspectives on local and global events. Likewise, the Octocopter, developed by Amazon books, explores the idea of using drones for the distribution of books. This is a contrast to Dutch artist Jansen's Orvillocopter, which has become a transmogrified ludic cat turned into a flying drone, after being hit by a car. These diverse examples redefine 'mobile' devices, which have 'social' dimensions. They are also part of a new 'digital age psychogeography' that includes other modes such as personal video clips distributed as video stamps™ using QR codes. To understand these various trends this research looks to notions of place, territory and community, drawing partly on locative media ideas introduced by Lemos. It also explores the similarities with shifts in urban wanderings that began in France with the Flâneur and crossed over into English literature in the 18th century, via the works of Blake, Defoe and others.

These works captured the imagined resonances of London, which over time shifted, as the literature and films of the 20th and 21st century revised psychogeography via Ballard, Self, Keillor and many others. From this background this research introduces a new discussion on 'digital age psychogeography', with various features identified in digital and social media, such as Facebook's Check-in feature and drones that require flight paths. And, like drones, checking- into map your location in social media depends on coordinates. This cluster of digital communication modes have evolved in wider contexts of maps, location and plotting techniques, visualization, as well as film concepts like the 'gaze'. The latter now applies digital media now roving across Mars, rural Australia, downtown Melbourne and many other universes. These digital communication modes will continue to serve functional and ludic possibilities, wherever urban and rural wanderings bypass skyscrapers, paddocks, exoplanets or a suburban underground. The notions of space, place and mobility require discernments about the physical journeys and applications in the 21st century as well as clarity of the mental travels and abstractions before us. Such discernments can ultimately inform future civic engagements for Australia. Against this background, this research suggests several new research projects to better understand the 'digital age of psychogeography'.

15:20- 16:50 **Sessions**

3.1: Language & Text III

Location: **Case Study Room**

Presentations

Navigating the 18th-Century Republic of Letters: Sequence Alignment as a Method for Scalable Reading in Large Text Corpora

Glenn H Roe

Digital Humanities Hub, Australian National University

The promise of "big data" approaches to humanities research has led in recent years to a host of innovative computational and algorithmic methods for exploring the growing digital human record. These techniques—from data mining to distant reading—offer students and scholars

new perspectives on otherwise intractable data sets; 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. This paper aims at addressing these concerns through the use of a text mining software package that applies sequence alignment algorithms—borrowed from the field of bio-informatics—to large-scale heterogeneous data sets in order to identify varying degrees of “intertextual” relationships.

The software, named “PAIR: Pairwise Alignment for Intertextual Relations” [Horton et al., 2010], was developed collaboratively by the University of Chicago’s ARTFL Project, and is currently being tested over several large literary and historical data sets in both Chicago and at the University of Oxford. In creating PAIR, we attempted to adapt existing techniques to suit the particular needs of humanities scholarship. Many algorithms such as BLAST [Altschul et al., 1990] exist for identifying duplicated DNA, for example, but these tend to emphasize speed over completeness. In text analysis, our corpora are small enough and our interest deep enough that we can emphasize retrieving as many hits as possible, while at the same time returning results that are of maximal interest, avoiding a preponderance of banal commonplaces and other formulaic phrases.

To achieve this, the PAIR system first indexes documents by breaking them into overlapping sequences of words, or “n-grams”, and then creates a database of the occurrences of each n-gram in a given corpus. This index allows for the discovery of text reuse by looking for occurrences of the same word sequences shared between documents or within different parts of the same document. Through various tuning parameters, this technique can be made flexible enough to find sequence matches with minor differences in word order, missing words, orthographic variations, misrecognized characters, and other textual variants. The first part of this paper will thus examine the philosophy behind the PAIR approach, the algorithmic design, its implementation as an open-source Perl module, and its application to a variety of tasks relevant to humanities research.

Building upon this methodological base, I will then present my current work using PAIR to explore the circulation of ideas and texts in the 18th-century “Republic of Letters”, and in particular, through the more than 60,000 letters contained in the University of Oxford’s *Electronic Enlightenment* database. The unique epistemological status of correspondence collections allows us to explore a more expansive area of “intertextuality”—one that incorporates recent work on social networks and communication circuits, historical media studies, and, more generally, literary, intellectual, and book history of the Enlightenment period.

By reintroducing this expanded notion of (inter)textuality into algorithmic and data-driven methods, we can perhaps bridge the gap between distant and close readings, through an intermediary mode of scholarship I term “scalable” reading, to borrow a term from Martin Mueller. I will further argue that the digital humanities as a field is perhaps uniquely poised to function as an interdisciplinary locus for just this sort of scalable textual analysis, moving between the distant and the close, the macro and micro, by way of a digitally-assisted “meso-analysis” informed both by new computational methods and the centuries-old tradition of textual scholarship.