## Introducing DH 2010

<table>
<thead>
<tr>
<th>Journal:</th>
<th>Literary and Linguistic Computing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID:</td>
<td>Draft</td>
</tr>
<tr>
<td>Manuscript Type:</td>
<td>Editorial</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>n/a</td>
</tr>
<tr>
<td>Complete List of Authors:</td>
<td>Nerbonne, John; University of Groningen, Humanities Computing Nowviskie, Bethany; University of Virginia Library, Scholars' Lab Spence, Paul; King's College London, Digital Humanities Vetch, Paul; King's College London, Digital Humanities</td>
</tr>
<tr>
<td>Keywords:</td>
<td>Digital Humanities, DH 2010, selected conference papers</td>
</tr>
</tbody>
</table>
Introduction

The Digital Humanities 2010 (DH 2010) conference took place from July 7th through 10th at King’s College London, where it was hosted by the Department of Digital Humanities (then Centre for Computing in the Humanities) and the Centre for e-Research, with the support of the School of Arts and Humanities, Information Services and Systems, and the Principal, Professor Rick Trainor. There were six satellite workshops, ten multi-speaker panels of ninety minutes each, eighty paper presentations, and twenty-three poster presentations. Over 300 scholars and students registered to participate in the conference.

DH 2010 was a special occasion for many reasons. The Busa award was presented to Joseph Raben, emeritus professor of English at Queens College of the City University of New York (CUNY), *inter alia* in recognition of his founding the journal *Computers and the Humanities* in 1966 and the Association for Computers and the Humanities in 1978. Raben delivered a provocative lecture suggesting the computer will ultimately change not just humanities scholarship, but also the academic institutions it is part of. The Fortier prize, named after the late Canadian specialist on French literature and humanities computing, to be given to the best presentation by a young scholar at the conference, was awarded for the first time. Maceij Eder, Kraków, received the prize for his work on non-traditional authorship attribution. One of Eder’s papers is included in this special issue. Masahiro Hori, Osamu Imahayashi, Tomoji Tabata, and Miyuki Nishio received the award for the best poster for ‘The Dickens Lexicon and its practical use for linguistic research.’

DH 2010 was remarkable in other ways, as well. King’s College London has long been a center for humanities computing world wide, and the local organizers aimed to build on the success of recent conferences, with a program which combined
For Peer Review

a traditional academic strand with a performance strand, reflecting the conference
theme of ‘Cultural expression, old and new’. London is an exhilarating city to visit,
and participants spotted each other at different sites throughout the city. Finally, DH
2010 was Harold Short’s last conference as chair of the Association for Literary and
Linguistic Computing (ALLC), a position he held for over ten years. It was fitting
that the ALLC, in the person of its incoming chair, Lisa Lena Opas-Hänninen, could
thank Harold for all he has done at King’s, his home institution, and also for the
profession. The ALLC owes a great debt to Harold’s energy and good judgment.

The international program committee (IPC) broke with traditions in recent DH
conferences, which have featured prominent researchers from neighboring fields as
plenary speakers. These speakers have often presented fascinating lectures, but the
DH 2010 IPC felt that the field has matured enough for it to be worthwhile to invite a
“core” digital humanities researcher to address the profession and identify where she
sees key issues and opportunities. Since the IPC also sought a younger practitioner in
view of the increasing numbers of younger researchers at the DH conferences, we
were pleased that Melissa Terras of University College London, graciously accepted.
We are also pleased to publish the lecture among the papers below.

Academic conferences are “stone soup”\(^1\) affairs that rely on voluntary
contributions, in particular, contributions of labor. DH 2010 was fortunate in being
able to count on an excellent IPC, including Elisabeth Burr (Leipzig), Richard
Cunningham (Acadia), Jan-Christoph Meister (Hamburg), Elli Mylonas (Brown),
Brent Nelson (Saskatchewan), Jan Rybicki (Kraków), and John Walsh (Indiana).
John Nerbonne and Bethany Nowviskie served as chair and vice-chair, respectively.
Harold Short, Paul Spence and Paul Vetch led a particularly industrious local

---

\(^1\) It turns out that some don’t know the reference. See http://en.wikipedia.org/wiki/Stone_soup
organizing committee, which included Sheila Anderson, Tobias Blanke, Gabriel Bodard, John Bradley, Sarah Davenport, Mark Davies, Stuart Dunne, Mark Hedges, Lorna Hughes, Carrie van de Langenberg, John Lavagnino, Willard McCarty, Elena Pierazzo, Torsten Reimer, Helen Skundric and Simon Tanner.

Everyone who presented at the conference was invited to submit a paper to this special issue of Literary and Linguistic Computing, and we received twenty-two submissions, all of which were reviewed and evaluated by three anonymous referees, leading to the selection of twelve papers being published in this special issue.

We do not presume to suggest that the twelve papers included in this special issue are representative of all the buzzing, blooming confusion of approaches, techniques, technologies, experimental designs, theoretical frameworks and reflective perspectives that our conferences abound in. Nor did we aim at a representative selection. Instead we sought to identify some of the better and more promising work. Naturally it is up to you – our community – to judge how successful we were in that.

We introduce each paper briefly in turn.

Papers

Melissa Terras gave the closing plenary speech at the conference, a presentation titled “Present, Not Voting: Digital Humanities in the Panopticon”. Although her talk closed the 2010 conference, we use it to open this special issue of papers from the conference. Terras’ talk explored some of the key challenges facing the Digital Humanities as a field in the face of adverse economic and political circumstances. Her paper, presented in the same informal and direct style with which it was delivered, uses the Transcribe Bentham project to examine a wide range of theoretical and practical issues: including the potential for crowd-sourcing techniques to open up historical sources, differing views of scholarly ‘impact’ and the importance
of articulating the real impact of the DH, and the particular complexities in sustaining
a highly interdisciplinary field like DH in a manner which both ensures that its
outcomes are properly recognized by the Academy as a whole and at the same time
provides real openings for younger scholars.

Marcus Bingenheimer, Jen-Jou Hung and Simon Wiles present a methodology
for generating social network visualizations from TEI-encoded textual materials of
great importance to Chinese Buddhist history. Their research, which has been applied
to a collection of Buddhist monk biographies, aims to demonstrate how creating GIS-
like visualizations of a text encoded for spatio-temporal analysis can enable scholars
to make inferences regarding social connections, an approach theoretically extensible
to any body of historical textual material with appropriate markup. The research
makes use of the PREFUSE toolkit and depends on the concept of a nexus-point,
simple event-based information in the text which connects people, places and time.

Claire Brierley and Eric Atwell introduce proPOSEL in their paper, a system
for phrase boundary detection. The authors predict how readers would normally
divide a written text into pronounceable phrases. ProPOSEL uses not only
punctuation and syntactic features (as in earlier models), but also features of vowel
pronunciation, which it infers from the written text. ProPOSEL’s pronunciation
features are unlikely to vary and therefore remain useful domain-independently. The
authors conduct a number of corpus studies to demonstrate that vowel complexity and
the position of phrase breaks is not independent, and they speculate that internal
prosody informs “chunking” in silent reading.

Christopher Forstall, Sarah Jacobson and Walter Scheirer describe their work
to apply and extend authorship attribution techniques to test the theory voiced by
various literary scholars that Paul the Deacon's Angustae Vitae might contain
borrowings from, and allusions to, Catullus. The literary implications of this are
highly significant given that Catullus’ work is commonly held to have remained
undiscovered at this time. The authors apply a one-class SVM approach to three
distinct sets of features which they argue could betoken intertextuality – phonetic,
dictional, and metrical. Their analysis provides objective evidence for the posited
intertextual relationship, and the authors report that the technology is now being
incorporated into a web-based allusion detection tool.

In "Visual GISting," Ian Gregory and Andrew Hardie demonstrate an
approach to analyzing the geographies within texts that integrates methodologies from
corpus linguistics and geographical information systems (GIS). The paper
summarizes the methods and aims of both fields. It then demonstrates the ways in
which an acknowledged lack of uptake in GIS among humanities scholars working in
text-based modes can be overcome using data-mining and corpus linguistics. Gregory
and Hardie describe the use of part-of-speech tagging and gazetteer comparison to
extract place names from a corpus, which can then be visualized and analyzed using a
GIS. The method is extended to the semantics of place-names, connecting proper
nouns with concepts related to war or historical economies, for example, and
visualizing these clusters geographically.

Jan Rybicki and Maceij Eder’s paper “Deeper Delta across Genres and
Languages: Do we really need the most Frequent Words?” presents a new step in
non-traditional authorship attribution. Rybicki and Eder systematically vary the
number and the frequency band of the words they use as features in authorship
attribution, initially using just the fifty most frequent words, then the next fifty in the
frequency list, etc. They also vary the number of words used, going from fifty to one
hundred, etc., and finally they suggest “heat maps” as an insightful visualization of
results. They are able to show that languages and genres differ in the ideal size
needed for authorship attribution and in the frequency bands of the words used as
features.

Philip Sabin offers an analysis of the surprisingly popular phenomenon of
‘manual’, offline war games (often based on printed maps), noting the thriving
subculture on the internet, where many such games are sold or freely distributed.

Sabin argues that, although impressive and useful in certain situations, computerized
models are costly to develop well, and can result in inaccurate simulation outcomes
(especially when recreating combatants’ behavior). On the other hand manual war
games have the potential to encourage much more analytical thinking in their users by
drawing them into the logic of the simulation, and they also encourage creativity by
making simulation design a more accessible possibility. As such, manual war games
have notable pedagogical value.

Maxime Sainte-Marie, Jean-Guy Meunier, Nicolas Payette and Jean-François
Chartier apply a computational linguistics algorithm for conceptual analysis
developed at the Laboratoire D’Analyse Cognitive de L’information in an attempt to
“read Darwin between the lines”. They explore the famous absence of the word
‘evolution’ (and cognates) from The Origin of Species, and examine the historical
development of the word, focusing on those senses which were routinely employed
about the time when Darwin was writing. Using the evidence gleaned from their
conceptual analysis, they argue that although the word ‘evolution’ itself is not
present, the modern sense of the word (i.e. ‘the development of species’, rather than
the contemporary sense of ‘embryological development’) prevails in different phrases
and word combinations.
In "A Tale of Two Cities," Lynne Siemens and her collaborators address similarities and differences in attitudes toward collaboration among members of the digital libraries and DH communities, who are now working jointly with increasing frequency. Their study contextualizes teamwork within both communities and with reference to solitary-scholar modes of production common in the academy. Survey and interview research contributed to the study and its authors conclude that collaborative efforts across the two communities are generally healthy and likely to improve. They distill typical attitudes toward team-based efforts among digital humanists and librarians, offer some characterizations of the two groups, and conclude with implications for future practice.

Ségolène Tarte's paper on "Digitizing the Act of Papyrological Interpretation" describes an eSAD (e-Science and Ancient Documents project) design philosophy for building digital tools to aid in the interpretation of incised documents that are difficult to read. The expert strategies of mimesis, as in tracing, and of evaluation of evidence toward hypotheses as is seen in crossword puzzle-solving are taken on board in an approach that views digitization as involving both sampling and interpretation and takes into account desired and time-tested interpretive practices of scholars with regard to ambiguity. The overall goal of the software is to allow researchers to avoid "spurious exactitude" in their readings, while embracing what Tarte calls "genuine uncertainty."

In “Finding Stories in the Archive through Paragraph Alignment”, Weijia Xu and Maria Esteva explore the potential of a novel paragraph level alignment technique as a means of identifying similar content within large, unorganized archives of data – as, for example, the sorts of procedural documents that accrue over time on shared network drives in most big organizations. They demonstrate the limitations of
traditional cosine similarity methodologies (which operate globally, i.e. on all the words in a document) in this situation, and evaluate their own ‘localized’ approach, which they successfully use to achieve better search precision. They observe that their approach may have considerable promise for similarity searching of other difficult corpora, such as email archives.

Amélie Zoellner-Weber’s paper “Text Encoding and Ontology – Enlarging an Ontology by Semi-Automatic Generated Instances” presents a framework which allows users to encode literary features and then semi-automatically match them to a given domain ontology. Tested using an extract from the novel *Melmoth the Wanderer* by Charles Robert Maturin, the application described allows a user to model literary phenomena based on their own reading of a text, to store statements about sections of the text during the reading process and to then formalise statements within an ontology.

John Nerbonne
Center for Language and Cognition
University of Groningen

Bethany Nowviskie
Scholars’ Lab
University of Virginia Library

Paul Spence
Department of Digital Humanities (DDH)
King’s College London

Paul Vetch

Department of Digital Humanities (DDH)

King’s College London