



**Introducing DH 2010**

Journal:	<i>Literary and Linguistic Computing</i>
Manuscript ID:	Draft
Manuscript Type:	Editorial
Date Submitted by the Author:	n/a
Complete List of Authors:	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
Keywords:	Digital Humanities, DH 2010, selected conference papers

SCHOLARONE™  
Manuscripts

## Introduction

The Digital Humanities 2010 (DH 2010) conference took place from July 7<sup>th</sup> through 10<sup>th</sup> 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

1  
2  
3 a traditional academic strand with a performance strand, reflecting the conference  
4 theme of ‘Cultural expression, old and new’. London is an exhilarating city to visit,  
5  
6 and participants spotted each other at different sites throughout the city. Finally, DH  
7  
8 2010 was Harold Short’s last conference as chair of the Association for Literary and  
9  
10 Linguistic Computing (ALLC), a position he held for over ten years. It was fitting  
11  
12 that the ALLC, in the person of its incoming chair, Lisa Lena Opas-Hänninen, could  
13  
14 thank Harold for all he has done at King’s, his home institution, and also for the  
15  
16 profession. The ALLC owes a great debt to Harold’s energy and good judgment.  
17  
18  
19  
20  
21

22 The international program committee (IPC) broke with traditions in recent DH  
23  
24 conferences, which have featured prominent researchers from neighboring fields as  
25  
26 plenary speakers. These speakers have often presented fascinating lectures, but the  
27  
28 DH 2010 IPC felt that the field has matured enough for it to be worthwhile to invite a  
29  
30 “core” digital humanities researcher to address the profession and identify where she  
31  
32 sees key issues and opportunities. Since the IPC also sought a younger practitioner in  
33  
34 view of the increasing numbers of younger researchers at the DH conferences, we  
35  
36 were pleased that Melissa Terras of University College London, graciously accepted.  
37  
38 We are also pleased to publish the lecture among the papers below.  
39  
40  
41  
42

43 Academic conferences are “stone soup”<sup>1</sup> affairs that rely on voluntary  
44  
45 contributions, in particular, contributions of labor. DH 2010 was fortunate in being  
46  
47 able to count on an excellent IPC, including Elisabeth Burr (Leipzig), Richard  
48  
49 Cunningham (Acadia), Jan-Christoph Meister (Hamburg), Elli Mylonas (Brown),  
50  
51 Brent Nelson (Saskatchewan), Jan Rybicki (Kraków), and John Walsh (Indiana).  
52  
53 John Nerbonne and Bethany Nowviskie served as chair and vice-chair, respectively.  
54  
55 Harold Short, Paul Spence and Paul Vetch led a particularly industrious local  
56  
57  
58  
59  
60

---

<sup>1</sup> It turns out that some don’t know the reference. See [http://en.wikipedia.org/wiki/Stone\\_soup](http://en.wikipedia.org/wiki/Stone_soup)

1  
2  
3 organizing committee, which included Sheila Anderson, Tobias Blanke, Gabriel  
4  
5 Bodard, John Bradley, Sarah Davenport, Mark Davies, Stuart Dunne, Mark Hedges,  
6  
7  
8 Lorna Hughes, Carrie van de Langenberg, John Lavagnino, Willard McCarty, Elena  
9  
10 Pierazzo, Torsten Reimer, Helen Skundric and Simon Tanner.  
11

12  
13 Everyone who presented at the conference was invited to submit a paper to  
14  
15 this special issue of Literary and Linguistic Computing, and we received twenty-two  
16  
17 submissions, all of which were reviewed and evaluated by three anonymous referees,  
18  
19 leading to the selection of twelve papers being published in this special issue.  
20  
21

22  
23 We do not presume to suggest that the twelve papers included in this special  
24  
25 issue are representative of all the buzzing, blooming confusion of approaches,  
26  
27 techniques, technologies, experimental designs, theoretical frameworks and reflective  
28  
29 perspectives that our conferences abound in. Nor did we aim at a representative  
30  
31 selection. Instead we sought to identify some of the better and more promising work.  
32  
33 Naturally it is up to you – our community – to judge how successful we were in that.  
34  
35 We introduce each paper briefly in turn.  
36  
37

### 38 39 **Papers**

40  
41 Melissa Terras gave the closing plenary speech at the conference, a  
42  
43 presentation titled “Present, Not Voting: Digital Humanities in the Panopticon”.  
44  
45 Although her talk closed the 2010 conference, we use it to open this special issue of  
46  
47 papers from the conference. Terras’ talk explored some of the key challenges facing  
48  
49 the Digital Humanities as a field in the face of adverse economic and political  
50  
51 circumstances. Her paper, presented in the same informal and direct style with which  
52  
53 it was delivered, uses the *Transcribe Bentham* project to examine a wide range of  
54  
55 theoretical and practical issues: including the potential for crowd-sourcing techniques  
56  
57 to open up historical sources, differing views of scholarly ‘impact’ and the importance  
58  
59  
60

1  
2  
3 of articulating the real impact of the DH, and the particular complexities in sustaining  
4 a highly interdisciplinary field like DH in a manner which both ensures that its  
5 outcomes are properly recognized by the Academy as a whole and at the same time  
6 provides real openings for younger scholars.  
7  
8  
9  
10  
11

12        Marcus Bingenheimer, Jen-Jou Hung and Simon Wiles present a methodology  
13 for generating social network visualizations from TEI-encoded textual materials of  
14 great importance to Chinese Buddhist history. Their research, which has been applied  
15 to a collection of Buddhist monk biographies, aims to demonstrate how creating GIS-  
16 like visualizations of a text encoded for spatio-temporal analysis can enable scholars  
17 to make inferences regarding social connections, an approach theoretically extensible  
18 to any body of historical textual material with appropriate markup. The research  
19 makes use of the PREFUSE toolkit and depends on the concept of a *nexus-point*,  
20 simple event-based information in the text which connects people, places and time.  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33

34        Claire Brierley and Eric Atwell introduce proPOSEL in their paper, a system  
35 for phrase boundary detection. The authors predict how readers would normally  
36 divide a written text into pronounceable phrases. ProPOSEL uses not only  
37 punctuation and syntactic features (as in earlier models), but also features of vowel  
38 pronunciation, which it infers from the written text. ProPOSEL's pronunciation  
39 features are unlikely to vary and therefore remain useful domain-independently. The  
40 authors conduct a number of corpus studies to demonstrate that vowel complexity and  
41 the position of phrase breaks is not independent, and they speculate that internal  
42 prosody informs "chunking" in silent reading.  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54

55        Christopher Forstall, Sarah Jacobson and Walter Scheirer describe their work  
56 to apply and extend authorship attribution techniques to test the theory voiced by  
57 various literary scholars that Paul the Deacon's *Angustae Vitae* might contain  
58  
59  
60

1  
2  
3 borrowings from, and allusions to, Catallus. The literary implications of this are  
4 highly significant given that Catullus' work is commonly held to have remained  
5 undiscovered at this time. The authors apply a one-class SVM approach to three  
6 distinct sets of features which they argue could betoken intertextuality – phonetic,  
7 dictional, and metrical. Their analysis provides objective evidence for the posited  
8 intertextual relationship, and the authors report that the technology is now being  
9 incorporated into a web-based allusion detection tool.

10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20 In "Visual GISTing," Ian Gregory and Andrew Hardie demonstrate an  
21 approach to analyzing the geographies within texts that integrates methodologies from  
22 corpus linguistics and geographical information systems (GIS). The paper  
23 summarizes the methods and aims of both fields. It then demonstrates the ways in  
24 which an acknowledged lack of uptake in GIS among humanities scholars working in  
25 text-based modes can be overcome using data-mining and corpus linguistics. Gregory  
26 and Hardie describe the use of part-of-speech tagging and gazetteer comparison to  
27 extract place names from a corpus, which can then be visualized and analyzed using a  
28 GIS. The method is extended to the semantics of place-names, connecting proper  
29 nouns with concepts related to war or historical economies, for example, and  
30 visualizing these clusters geographically.

31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46 Jan Rybicki and Maceij Eder's paper "Deeper Delta across Genres and  
47 Languages: Do we really need the most Frequent Words?" presents a new step in  
48 non-traditional authorship attribution. Rybicki and Eder systematically vary the  
49 number and the frequency band of the words they use as features in authorship  
50 attribution, initially using just the fifty most frequent words, then the next fifty in the  
51 frequency list, etc. They also vary the number of words used, going from fifty to one  
52 hundred, etc., and finally they suggest "heat maps" as an insightful visualization of

1  
2  
3 results. They are able to show that languages and genres differ in the ideal size  
4  
5 needed for authorship attribution and in the frequency bands of the words used as  
6  
7 features.  
8  
9

10 Philip Sabin offers an analysis of the surprisingly popular phenomenon of  
11  
12 ‘manual’, offline war games (often based on printed maps), noting the thriving  
13  
14 subculture on the internet, where many such games are sold or freely distributed.  
15  
16

17 Sabin argues that, although impressive and useful in certain situations, computerized  
18  
19 models are costly to develop well, and can result in inaccurate simulation outcomes  
20  
21 (especially when recreating combatants' behavior). On the other hand manual war  
22  
23 games have the potential to encourage much more analytical thinking in their users by  
24  
25 drawing them into the logic of the simulation, and they also encourage creativity by  
26  
27 making simulation design a more accessible possibility. As such, manual war games  
28  
29 have notable pedagogical value.  
30  
31  
32  
33

34 Maxime Sainte-Marie, Jean-Guy Meunier, Nicolas Payette and Jean-François  
35  
36 Chartier apply a computational linguistics algorithm for conceptual analysis  
37  
38 developed at the *Laboratoire D'Analyse Cognitive de L'information* in an attempt to  
39  
40 “read Darwin between the lines”. They explore the famous absence of the word  
41  
42 ‘evolution’ (and cognates) from *The Origin of Species*, and examine the historical  
43  
44 development of the word, focusing on those senses which were routinely employed  
45  
46 about the time when Darwin was writing. Using the evidence gleaned from their  
47  
48 conceptual analysis, they argue that although the word ‘evolution’ itself is not  
49  
50 present, the modern sense of the word (i.e. ‘the development of species’, rather than  
51  
52 the contemporary sense of ‘embryological development’) prevails in different phrases  
53  
54 and word combinations.  
55  
56  
57  
58  
59  
60

1  
2  
3 In "A Tale of Two Cities," Lynne Siemens and her collaborators address  
4  
5 similarities and differences in attitudes toward collaboration among members of the  
6  
7 digital libraries and DH communities, who are now working jointly with increasing  
8  
9 frequency. Their study contextualizes teamwork within both communities and with  
10  
11 reference to solitary-scholar modes of production common in the academy. Survey  
12  
13 and interview research contributed to the study and its authors conclude that  
14  
15 collaborative efforts across the two communities are generally healthy and likely to  
16  
17 improve. They distill typical attitudes toward team-based efforts among digital  
18  
19 humanists and librarians, offer some characterizations of the two groups, and  
20  
21 conclude with implications for future practice.  
22  
23  
24  
25  
26

27 Ségolène Tarte's paper on "Digitizing the Act of Papyrological Interpretation"  
28  
29 describes an eSAD (e-Science and Ancient Documents project) design philosophy for  
30  
31 building digital tools to aid in the interpretation of incised documents that are difficult  
32  
33 to read. The expert strategies of mimesis, as in tracing, and of evaluation of evidence  
34  
35 toward hypotheses as is seen in crossword puzzle-solving are taken on board in an  
36  
37 approach that views digitization as involving both sampling and interpretation and  
38  
39 takes into account desired and time-tested interpretive practices of scholars with  
40  
41 regard to ambiguity. The overall goal of the software is to allow researchers to avoid  
42  
43 "spurious exactitude" in their readings, while embracing what Tarte calls "genuine  
44  
45 uncertainty."  
46  
47  
48  
49

50  
51 In "Finding Stories in the Archive through Paragraph Alignment", Weijia Xu  
52  
53 and Maria Esteva explore the potential of a novel paragraph level alignment  
54  
55 technique as a means of identifying similar content within large, unorganized archives  
56  
57 of data – as, for example, the sorts of procedural documents that accrue over time on  
58  
59 shared network drives in most big organizations. They demonstrate the limitations of  
60



1  
2  
3 traditional cosine similarity methodologies (which operate globally, i.e. on all the  
4 words in a document) in this situation, and evaluate their own ‘localized’ approach,  
5  
6 which they successfully use to achieve better search precision. They observe that  
7  
8 their approach may have considerable promise for similarity searching of other  
9  
10  
11  
12  
13 difficult corpora, such as email archives.

14  
15 Amélie Zoellner-Weber’s paper “Text Encoding and Ontology – Enlarging an  
16  
17 Ontology by Semi-Automatic Generated Instances” presents a framework which  
18  
19 allows users to encode literary features and then semi-automatically match them to a  
20  
21 given domain ontology. Tested using an extract from the novel *Melmoth the*  
22  
23 *Wanderer* by Charles Robert Maturin, the application described allows a user to  
24  
25 model literary phenomena based on their own reading of a text, to store statements  
26  
27 about sections of the text during the reading process and to then formalise statements  
28  
29 within an ontology.  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

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)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

King's College London

Paul Vetch

Department of Digital Humanities (DDH)

King's College London

For Peer Review