EACL 2014

**14th Conference of the European Chapter
of the Association for Computational Linguistics**

**Proceedings of the 3rd Workshop
on Computational Linguistics for Literature (CLfL)**

April 27, 2014
Gothenburg, Sweden

Order copies of this and other ACL proceedings from:

Preface

Welcome to the third edition of the workshop on Computational Linguistics for Literature. What started out two years ago as a small affair with an unorthodox title seems to be shaping into a modestly sized but vibrant research community.

Thanks to the effort of the authors and the program committee, April 27, 2014 promises to be an interesting day.

This year's workshop boasts the rich pickings of papers focussed on creating character representations from text and on applications of such representations. Agarwal et al. present a system for extracting social networks from movie scripts. Coll Ardanuy and Sporleder propose a method of finding similar novels using social networks as the underlying representation. Bullard and Alm work on identifying social information about characters from dialogues in a corpus of plays. Taking this a step further, Iosif and Mishra describe an integrated system for identifying and classifying characters in children's stories based on direct and indirect speech.

Taking a complementary point of view, the position paper by Levison and Lessard builds upon their previous work and proposes a graph-based representation for the temporal structure of narratives. The paper by Zemánek and Milička describes a diachronic study of Arabic literature, tracing the influence of certain treatises across centuries.

In the categories all of their own are two more papers. Davis and Mohammad break new ground with a paper on generating music from literature; you can listen to a one-minute musical summary of a novel. Mike Kestemont presents a position paper on authorship attribution, a topic which was quite popular in the 2012 and 2013 editions of the workshop.

Last but by no means least, we are delighted to have two invited speakers this year. An artist and researcher María Mencía from Kingston University in London will talk about electronic poetry. Jan Christoph Meister from University of Hamburg will tell us about the connections between narratology and computation.

We look forward to seeing you in Göteborg.

Anna Feldman, Anna Kazantseva, Stan Szpakowicz

Workshop organizers

Organizers:

Anna Feldman (Montclair State University)

Anna Kazantseva (University of Ottawa)

Stan Szpakowicz (University of Ottawa)

Program Committee:

Cecilia Ovesdotter Alm (Rochester Institute of Technology)

Apoorv Agarwal (Columbia University)

Julian Brooke (University of Toronto)

Hal Daumé III (University of Maryland)

David Elson (Google)

Micha Elsner (Ohio State University)

Mark Finlayson (MIT)

Pablo Gervás (Universidad Complutense de Madrid)

Roxana Girju (University of Illinois at Urbana-Champaign)

Catherine Havasi (MIT Media Lab)

Jerry Hobbs (University of Southern California)

Justine Kao (Stanford University)

Victor Kuperman (McMaster University)

Inderjeet Mani (Yahoo!)

Rada Mihalcea (University of Michigan)

David Mimno (Cornell University)

Saif Mohammad (National Research Council Canada)

Ekaterina Neklyudova (McMaster University)

Vivi Nastase (FBK Trento)

Owen Rambow (Columbia University)

Michaela Regneri (Saarland University)

Reid Swanson (University of California, Santa Cruz)

Rob Voigt (Stanford University)

Marilyn Walker (University of California, Santa Cruz)

Janice Wiebe (University of Pittsburgh)

Bei Yu (Syracuse University)

Table of contents

Workshop program

Sunday, April 27, 2014

Session I

9:00–9:05   Welcome

9:05–10:00   *Language-Art Digital Poetics: An exploration of digital textualities in the production of artistic research* (invited talk)
María Mencía

10:00–10:30 *Generating Music from Literature*
Hannah Davis and Saif Mohammad

10:30–11:00 Coffee break

Session II

11:00–11:30 *Computational analysis to explore authors' depiction of characters*
Joseph Bullard and Cecilia Ovesdotter Alm

11:30–12:00 *Quotations, Relevance and Time Depth: Medieval Arabic Literature in Grids and Networks*
Petr Zemánek and Jiří Milička

12:00–12:20 *Time after Time: Representing Time in Literary Texts*
Michael Levison and Greg Lessard

12:20–12:30 Free-for-all

12:30–14:00 Lunch break

María Mencía is an artist-researcher and Senior Lecturer in New Media Theory and Digital Media Practice in the School of Performance and Screen Studies at Kingston University, UK. She holds a PhD in Digital Poetics and Digital Art by the University of the Arts, London. She studied English Philology at the Complutense University in Madrid, Fine Art and History and Theory of Art at the University of the Arts London.

Mencía's practice-based research in language, art and technology draws from different cultural, social, artistic and literary traditions such as: linguistics, fine art, film, visual, concrete and sound poetry, with digital poetics, electronic writing, and new media art theories and practices. Her practice includes interactive installations, performances, net.art, sound-generated poems and interactive generative narratives.

http://www.mariamencia.com/

Sunday, April 27, 2014 (continued)

Session III

14:00–15:00   *Between the Digital Scylla and the Hermeneutic Charybdis: Digital Humanities*
(invited talk)
Jan Christoph Meister

15:00–15:30   *Structure-based Clustering of Novels*
Mariona Coll Ardanuy and Caroline Sporleder

15:30–16:00   Coffee break

Session IV

16:00–16:30   *From Speaker Identification to Affective Analysis: A Multi-Step System for Analyzing Children's Stories*
Elias Iosif and Taniya Mishra

16:30–17:00   *Parsing Screenplays for Extracting Social Networks from Movies*
Apoorv Agarwal, Sriramkumar Balasubramanian, Jiehan Zheng and Sarthak Dash

17:00–17:20   *Function Words in Authorship Attribution. From Black Magic to Theory?*
Mike Kestemont

17:20–17:30   Wrap-up

Jan Christoph Meister, a Professor of German Literature at the University of Hamburg, specializes in Narratology and Digital Humanities. He has published on narratological theory, humanities computing and various German authors, including Goethe, Gustav Meyrink and Leo Perutz. He currently serves as director of the Interdisciplinary Centre for Narratology at the University of Hamburg.

http://www.slm.uni-hamburg.de/ifg2/personal/jan-christoph-meister.html

Language-Art Digital Poetics: An exploration of digital textualities in the production of artistic research

María Mencía
Faculty of Arts and Social Sciences
Kingston University

Abstract

As an academic and an artist my practice-based research is at the intersection of language, art and digital technology. It explores the area of the in-between, the visual, the aural and the semantic. I am always interested in experimenting with the digital medium with the aim of engaging the reader/viewer/user in an experience of shifting 'in' and 'out' of language. This involves looking 'at' and looking 'through' transparent and abstract textualities and linguistic soundscapes. It draws from avant-garde poetics re-mediating concepts of reading and writing, exploring digital media grammars (voice activation, use of webcam, use of mouse, acts of revealing, triggering, cut and paste, dragging) for interactivity, aesthetics, engagement and meaning production. It is trans-disciplinary, bringing together different cultural, artistic and literary traditions such as: linguistics, fine art, visual, concrete and sound poetry, with digital poetics, electronic writing, and new media art theories and practices.

Between the Digital Scylla and the Hermeneutic Charybdis: Digital Humanities

Jan Christoph Meister
Department of Languages, Literatures and Media
Faculty of Arts
University of Hamburg

Abstract

The Digital Humanities are undoubtedly en vogue. Hardly any of the Humanities can nowadays ignore this methodological paradigm which affords us with vast resources of openly accessible digitized source material ranging from texts to multimedia files to 3D representations of material artifacts; with a plethora of digital research tools to explore and manipulate these primary resources; with virtual research environments that allow us to collaborate in real time across time and space; with exciting new forms of research publication and dissemination; and finally with big data approaches that have begun to balance the historically and qualitative oriented speculative concerns of the traditional Humanities with the more empirically oriented, descriptive research procedures of the Social Sciences.

To those who championed what was initially called 'Humanities Computing' since the late 1980s this development is of course rewarding. However, DH is not the first empiricist methodological paradigm whose advent the Humanities have witnessed during the latter part of the 20th century. Artificial Intelligence and Linguistics, to mention but two, also had their heyday – and in the 1970s many of us believed that these were to become the new master disciplines that would revolutionize the Humanities at large. It just didn't happen: Linguistics ranks as an equal again while AI has all but disappeared. DH, too, will experience its life cycle, and my personal prediction is that in ten years' time the term in its narrower sense of 'digital research methods in the humanities' will have become obsolete precisely because DH practices will by then have become business as usual for everyone.

Indeed, I think that there is something that is significantly more important than to engage in any such speculation about the future development of DH. One aspect that this methodological paradigm still lacks is a sufficiently developed critique of its own philosophical and epistemological fundamentals. If I were an opponent of DH this is the weak spot that I would attack relentlessly: how can a practice, a method, let alone a discipline that shies away from methodological self-reflection claim to be 'humanistic'?

Against this backdrop I would like to present some ideas on what I term the "DH-paradox": the inherent tension and, perhaps, even contradiction between the hermeneutic quest for holistic meaning (the German Sinn) - which, in Fregian terms, denotes more than just symbolic 'reference', but rather the coupling of reference and existential relevance that characterizes human symbolic systems and interactions – and the conceptual fundamental of the digital: the idea and premise that, whatever we want to study can be broken down into discrete, measurable atomic observations. The big and fascinating challenge for DH and for the Humanities, I believe, is how we can make this tension between the holistic and the atomic, the synthetic human experience and the analytic formalism of abstraction become productive from a Humanists perspective.