NAACL HLT 2015

The 2015 Conference of the
North American Chapter of the
Association for Computational Linguistics:
Human Language Technologies

Proceedings of the Fourth Workshop on Computational Linguistics for Literature

June 4, 2015
Denver, Colorado, USA

Order print-on-demand copies from:

Curran Associates
57 Morehouse Lane
Red Hook, New York 12571
USA
Tel: +1-845-758-0400
Fax: +1-845-758-2633
curran@proceedings.com

ISBN 978-1-941643-36-5

Preface

Welcome to the 4^{th} edition of the Workshop on Computational Linguistics for Literature. After the rounds in Montréal, Atlanta and Göteborg, we are pleased to see both the familiar and the new faces in Denver.

We are eager to hear what our invited speakers will tell us. Nick Montfort, a poet and a pioneer of digital arts and poetry, will open the day with a talk on the use of programming to foster exploration and fresh insights in the humanities. He suggests a new paradigm useful for people with little or no programming experience.

Matthew Jockers's work on macro-analysis of literature is well known and widely cited. He has published extensively on using digital analysis to view literature diachronically. Matthew will talk about his recent work on modelling the shape of stories via sentiment analysis.

This year's workshop will feature six regular talks and eight posters. If our past experience is any indication, we can expect a lively poster session. The topics of the 14 accepted papers are diverse and exciting.

Once again, there is a lot of interest in the computational analysis of poetry. Rodolfo Delmonte will present and demo SPARSAR, a system which analyzes and visualizes poems. Borja Navarro-Colorado will talk about his work on analyzing shape and meaning in the 16^{th} and 17^{th} century Spanish sonnets. Nina McCurdy, Vivek Srikumar & Miriah Meyer propose a formalism for analyzing sonic devices in poetry and describe an open-source implementation.

This year's workshop will witness a lot of work on parallel texts and on machine translation of literary data. Laurent Besacier & Lane Schwartz describe preliminary experiments with MT for the translation of literature. In a similar vein, Antonio Toral & Andy Way explore MT on literary data but between related languages. Fabienne Cap, Ina Rösiger & Jonas Kuhn explore how parallel editions of the same work can be used for literary analysis. Olga Scrivner & Sandra Kübler also look at parallel editions – in dealing with historical texts.

Several other papers cover various aspects of literary analysis through computation. Prashant Jayannavar, Apoorv Agarwal, Melody Ju & Owen Rambow consider social network analysis for the validation of literary theories. Andreas van Cranenburgh & Corina Koolen investigate what distinguishes literary novels from less literary ones. Dimitrios Kokkinakis, Ann Ighe & Mats Malm use computational analysis and leverage literature as a historical corpus in order to study typical vocations of women in the 19^{th} century Sweden. Markus Krug, Frank Puppe, Fotis Jannidis, Luisa Macharowsky, Isabella Reger & Lukas Weimar describe a coreference resolution system designed specifically with fiction in mind. Stefan Evert, Thomas Proisl, Thorsten Vitt, Christof Schöch, Fotis Jannidis & Steffen Pielström explain the success of Burrows's Delta in literary authorship attribution.

Last but not least, there are papers which do not fit into any other bucket. Marie Dubremetz & Joakim Nivre will tell us about automatic detection of a rare but elegant rhetorical device called *chiasmus*. Julian Brooke, Adam Hammond & Graeme Hirst describe a tool much needed in the community: GutenTag, a system for accessing Project Gutenberg as a corpus.

To be sure, there will be much to listen to, learn from and discuss for everybody with the slightest interest in either NLP or literature. We cannot wait for June 4 (-:).

This workshop would not have been possible without the hard work of our program committee. Many people on the PC have been with us from the beginning. Everyone offers in-depth, knowledgeable advice to both the authors and the organizers. Many thanks to you all! We would also like to acknowledge the generous support of the National Science Foundation (grant No. 1523285), which has allowed us to invite such interesting speakers.

We look forward to seeing you in Denver!

Anna F., Anna K., Stan and Corina

Nick Montfort (http://nickm.com/)

Short bio

Nick Montfort develops computational art and poetry, often collaboratively. He is on the faculty at MIT in CMS/Writing and is the principal of the naming firm Nomnym. Montfort wrote the books of poems *#!* and *Riddle & Bind*, co-wrote *2002: A Palindrome Story*, and developed more than forty digital projects including the collaborations *The Deletionist* and *Sea and Spar Between*. The MIT Press has published four of his collaborative and individual books: *The New Media Reader*, *Twisty Little Passages*, *Racing the Beam*, and *10 PRINT CHR$(205.5+RND(1)); : GOTO 10*, with *Exploratory Programming for the Arts and Humanities* coming soon.

Invited talk: *Exploratory Programming for Literary Work*

Abstract

We are fortunate to be at a stage when formal research projects, including substantial ones on a large scale, are bringing computation to bear on literary questions. While I participate in this style of research, in this talk I plan to discuss some different but valuable approaches to literature that use computation, approaches that are complementary. Specifically, I will discuss how smaller-scale and even ad hoc explorations can lead to new insights and suggest possibilities for more structured and larger-scale research. In doing so, I will explain my concept of exploratory programming, a style of programming that I find particularly valuable in my own practice and research and that I have worked to teach to students in the humanities, most of whom have no programming background. I am completing a book, *Exploratory Programming for the Arts and Humanities*, to be published by the MIT Press next year, which I hope will foster this type of programming. In my talk, I will provide some examples of how both generative approaches (developing system that produce literary language) and analytical approaches can be undertaken using exploratory programming, and will explain how these can inform one another. While some of this work has been presented in literary studies and computer science contexts, my examples will also include work presented in art and poetry contexts.

Matthew Jockers (http://www.matthewjockers.net/)

Short bio

Matthew L. Jockers is Associate Professor of English at the University of Nebraska, Faculty Fellow in the Center for Digital Research in the Humanities, Faculty Fellow in the Center for Great Plains Studies, and Director of the Nebraska Literary Lab. His books include *Macroanalysis: Digital Methods and Literary History* (University of Illinois, 2013) and *Text Analysis Using R for Students of Literature* (Springer, 2014). He has written articles on computational text analysis, authorship attribution, Irish and Irish-American literature, and he has co-authored several amicus briefs defending the fair and transformative use of digital text.

Invited talk: *The (not so) Simple Shape of Stories*

Abstract

In a now famous lecture, Kurt Vonnegut described what he called the "simple shapes of stories." His thesis was that we could understand the plot of novels and stories by tracking fluctuations in sentiment. He illustrated his thesis by drawing a grid in which the y-axis represented "good fortune" at the top and "ill fortune" at the bottom. The x-axis represented narrative time and moved from "beginning" at the left to "end" at the right. Using this grid, Vonnegut traced the shapes of several stories including what he called the "man in hole" and the "boy meets girl". At one point in the lecture, Vonnegut wonders why computers cannot be trained to reveal the simple shapes of stories. In this lecture, Matthew Jockers will describe his attempt to model the simple shapes of stories using methods from sentiment analysis and signal processing.

Program Committee

Apoorv Agarwal (Columbia University)
Cecilia Ovesdotter Alm (Rochester Institute of Technology)
David Bamman (Carnegie Melon University)
Peter Boot (Huygens Institute for Netherlands History)
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)
Amit Goyal (University of Maryland)
Catherine Havasi (MIT Media Lab)
Justine Kao (Stanford University)
David Mimno (Cornell University)
Saif Mohammad (National Research Council Canada)
Rebecca Passonneau (Columbia University)
Livia Polanyi (LDM Associates)
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)

Invited Speakers

Matthew Jockers (University of Nebraska)
Nick Montfort (MIT)

Organizers

Anna Feldman (Montclair State University)
Anna Kazantseva (University of Ottawa)
Stan Szpakowicz (University of Ottawa)
Corina Koolen (Universiteit van Amsterdam)

Table of Contents

Conference Program

Thursday, June 4, 2015

Session 1

8:57–9:00 Welcome

9:00–10:00 *Exploratory Programming for Literary Work* (invited talk)
Nick Montfort

10:00–10:30 *Tools for Digital Humanities: Enabling Access to the Old Occitan Romance of Flamenca*
Olga Scrivner and Sandra Kübler

Coffee break

Session 2

11:00–11:30 *RhymeDesign: A Tool for Analyzing Sonic Devices in Poetry*
Nina McCurdy, Vivek Srikumar and Miriah Meyer

11:30–12:00 *Rhetorical Figure Detection: the Case of Chiasmus*
Marie Dubremetz and Joakim Nivre

12:00–12:30 *Validating Literary Theories Using Automatic Social Network Extraction*
Prashant Jayannavar, Apoorv Agarwal, Melody Ju and Owen Rambow

Thursday, June 4, 2015 (continued)

Lunch break

Session 3

14:00–15:00 *The (not so) Simple Shape of Stories* (invited talk)
Matthew Jockers

15:00–15:30 Poster teaser talks

Coffee break

Session 4

16:00–16:30 Poster session

GutenTag: an NLP-driven Tool for Digital Humanities Research in the Project Gutenberg Corpus
Julian Brooke, Adam Hammond and Graeme Hirst

A Pilot Experiment on Exploiting Translations for Literary Studies on Kafka's "Verwandlung"
Fabienne Cap, Ina Rösiger and Jonas Kuhn

Identifying Literary Texts with Bigrams
Andreas van Cranenburgh and Corina Koolen

Visualizing Poetry with SPARSAR – Visual Maps from Poetic Content
Rodolfo Delmonte

Towards a better understanding of Burrows's Delta in literary authorship attribution
Stefan Evert, Thomas Proisl, Thorsten Vitt, Christof Schöch, Fotis Jannidis and Steffen Pielström

Gender-Based Vocation Identification in Swedish 19th Century Prose Fiction using Linguistic Patterns, NER and CRF Learning
Dimitrios Kokkinakis, Ann Ighe and Mats Malm

Thursday, June 4, 2015 (continued)

Rule-based Coreference Resolution in German Historic Novels
Markus Krug, Frank Puppe, Fotis Jannidis, Luisa Macharowsky, Isabella Reger and Lukas Weimar

A computational linguistic approach to Spanish Golden Age Sonnets: metrical and semantic aspects
Borja Navarro

Session 5

16:30–17:00 *Automated Translation of a Literary Work: A Pilot Study*
Laurent Besacier and Lane Schwartz

17:00–17:30 *Translating Literary Text between Related Languages using SMT*
Antonio Toral and Andy Way

17:30–17:33 Farewell