CLfL at NAACL 2013

The 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies

Proceedings of the Second Workshop on Computational Linguistics for Literature

June 14, 2013
Atlanta, GA, USA

209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-937284-47-3

A word from the organizers

Welcome to the second edition of our young but vibrant workshop on Computational Linguistics for Literature. We are thrilled to have been able to accept a pleasantly wide range of interesting papers on the computational treatment of literature. The ACL community is certainly embracing literature!

We want the workshop to bring together NLP researchers interested in literature and literary scholars on the quantitative edge of their field. We feel that those who "count words" for a living have something to offer to people who "read books" for a living, and *vice versa*. As Rauscher *et al.* (this volume) put it:

> It is hard for the computer scientist to imagine what research questions form the discourse in the humanities. In contrast to this, humanities scholars have a hard time imagining the possibilities and limitations of computer technology. . .

Most papers at this year's workshop touch upon the mutual benefits of interdisciplinary examination and the hurdles between computational methods and literary analysis. Two papers discuss such issues directly. Hammon *et al.* share their experience of combining literary analysis and computation in an annotation project. They emphasize the advantages of such collaboration. Boot discusses the importance of research into how literary works are perceived by their audiences and how a corpus of written "responses" can be a useful and interesting resource. This line of research, if further developed, may help gain insights into the role of the reader in the literary process – and help show the way toward modeling that role computationally.

Two papers look at similar problems: how NLP can be effective in exploring and comparing differences between genres, and in testing certain literary hypotheses. Rauscher *et al.* show that extended analysis of concordances helps gain literary insight; and Jautze *et al.* use syntax as the basis of informative stylometric analysis across genres.

Like last year, the computational treatment of poetry takes the central role at the workshop: five of ten papers! Voigt and Jurafsky perform a diachronic study of how the 20th-century political history of China has affected the country's poetic tradition. Asgari and Chappelier apply topic modeling to a corpus of Persian poems and demonstrate that their methodology can contribute to comparative literature studies. Almuhareb *et al.* work on distinguishing Arabic poems from prose, and develop a search engine for Arabic poetry. The other two papers deal with high-level topical analyses of poetry, and point out significant challenges in this task. Fournier describes a pilot study into topical segmentation of Coleridge's Kubla Khan; Brooke *et al.* build upon their previous work on topical segmentation of T. S. Eliot's The Waste Land in an attempt to automatically cluster its segments by their speakers.

At the symbolic end of the spectrum this year, Lessard and Levison explore the process of constructing a graphical representation of a story's event structure to examine the role of repetition. They find the directed acyclic graph (DAG) to be a formalism that captures the intersecting threads of time and action in the film *Groundhog Day*.

Finally, we are honored to host two distinguished speakers for a pair of invited talks. Livia Polanyi, Consulting Professor of Linguistics at Stanford University, has made longstanding contributions to the study of narrative, computational linguistics, discourse theory and related fields. Mark Riedl, Assistant Professor in the Georgia Tech School of Interactive Computing and director of its Entertainment Intelligence Lab, is an expert in the emerging field of interactive narrative and its relationship with the study of textual narrative.

It is already the second edition of our workshop, and yet we are still only just scratching the surface of what interesting computational and humanistic problems – and solutions – are found in the collaboration of computational linguistics and literary analysis… Enjoy!

Anna, David and Stan

Program Committee

Apoorv Agarwal, Columbia University
Cecilia Ovesdotter Alm, Rochester Institute of Technology
Nate Chambers, United States Naval Academy
Nicholas Dames, Columbia University
Anna Feldman, Montclair State University
Mark Finlayson, MIT
Pablo Gervás, Universidad Complutense de Madrid
Amit Goyal, University of Maryland
Catherine Havasi, MIT Media Lab
Jerry Hobbs, University of Southern California
Justine Kao, Stanford University
Kathy McKeown, Columbia University
Inderjeet Mani, Yahoo Labs!
Rada Mihalcea, University of North Texas
Saif Mohammad, National Research Council, Canada
Vivi Nastase, FBK Trento
Rebecca Passonneau, Columbia University
Livia Polanyi, Stanford University
Michaela Regneri, Saarland University
Reid Swanson, University of California, Santa Cruz
Marilyn Walker, University of California, Santa Cruz
Janyce Wiebe, University of Pittsburgh
Bei Yu, Syracuse University

Invited Speakers

Livia Polanyi, Stanford University
Mark Riedl, Georgia Tech

Organizers

David Elson, Google
Anna Kazantseva, University of Ottawa
Stan Szpakowicz, University of Ottawa

The papers

The schedule

9:00	Welcome
9:00-10:00	Invited talk 1 *Reflections on Verbal Art 40 years after* Livia Polanyi
10:00-10:30	*A Tale of Two Cultures: Bringing Literary Analysis and Computational Linguistics Together* Adam Hammond, Julian Brooke and Graeme Hirst
10:30-11:00	Coffee break
11:00-11:30	*Recognition of Classical Arabic Poems* Abdulrahman Almuhareb, Ibrahim Alkharashi, Lama Al Saud and Haya Altuwaijri
11:30-12:00	*Tradition and Modernity in 20th Century Chinese Poetry* Rob Voigt and Dan Jurafsky
12:00-12:30	*Linguistic Resources and Topic Models for the Analysis of Persian Poems* Ehsaneddin Asgari and Jean-Cedric Chappelier
12:30-14:00	Lunch break
14:00-15:00	Invited talk 2 *Intelligent Narrative Generation: From Cognition to Crowdsourcing* Mark Riedl
15:00-15:30	Poster teasers *The desirability of a corpus of online book responses* Peter Boot *Clustering Voices in The Waste Land* Julian Brooke, Graeme Hirst and Adam Hammond *An initial study of topical poetry segmentation* Chris Fournier *Groundhog DAG: Representing Semantic Repetition in Literary Narratives* Greg Lessard and Michael Levison *Exploring Cities in Crime: Significant Concordance and Co-occurrence in Quantitative Literary Analysis* Janneke Rauscher, Leonard Swiezinski, Martin Riedl and Chris Biemann
15:30-16:00	Coffee break
16:00-16:30	Poster session
16:30-17:00	*From high heels to weed attics: a syntactic investigation of chick lit and literature* Kim Jautze, Corina Koolen, Andreas van Cranenburgh and Hayco de Jong
17:00-17:30	An informal presentation *Identification of Speakers in Novels* Hua He, Denilson Barbosa and Greg Kondrak
17:30	Farewell

Invited speaker 1

Livia Polanyi joined Stanford University in 2012 as Consulting Professor of Linguistics after leaving Microsoft Corporation where she was a Principal Researcher at Bing working on applications of formal theories of discourse structure to problems in search. She also taught at the University of Amsterdam, Rice University and the University of Tel Aviv, and held scientist positions in computational linguistics at BBN Labs, Fuji-Xerox Palo Alto Labs and Powerset Corporation where she was the first member of the technical team. Professor Polanyi's research focusses on the structure of language above the sentence and she has published work in theoretical, socio and computational linguistics as well as in literary theory, anthropology, economics and political science. Currently she is working on extensions of formal concepts developed to account for discourse interpretability despite discontinuity to foundational problems in music, dance and conversation. She is also a poet.

Reflections on Verbal Art 40 years after

Abstract

Many years ago, a young girl who dared not call herself a "poet" sat alone at her desk day after day writing texts that she dared not call "poems". Each text, she knew, was a little theory about the nature of language. Once the text began, she simply did what was required. The artist is servant not mistress, she had learned. Creation is strictly carrying out what one is told. It was a lonely life. No one saw the words that she wrote. As a foot soldier in the army of those who wrote for the desk drawer, she talked to herself and grew impatient with what she had to say and so she decided one day to leave her desk and go out into the world beyond the window, to read instead of write and to listen to what others had to say. She wanted to understand what she had been doing day after day at her desk near the window because she knew that what had been happening there was that literature was being born – not great literature, probably not even good literature – and she wanted to know what this "literature" she was so busy serving might be. And so, she began to study the nature of language and to put off all consideration of what on earth that girl behind the window had been doing. She read and she studied and she learned and eventually she wrote and she explained and she taught and she left behind the questions about the nature of literature that had sent her out in the world to understand.

But the years have a way of catching up with everyone and as summers became winters and winters became summers again, the young girl became a young woman and then a woman no longer young and then again that young woman no longer young found herself alone at a desk writing texts the she dared not call "poems" but now as she wrote as the theories flowed out onto the page she had learned enough to understand what these texts were theories of and why and how the language of the first breath that would grow into a full text determined the possibilities of what that text could become. And so, in this talk, having been asked to talk to this workshop on computational approaches to literature I will share with you some speculations about the nature of Verbal Art that I have learned through study and practice in the years that separated the woman who will stand before you from that young girl who sat behind her desk near a window years and years ago and wrote down texts that she did not give the name that they had earned.

Invited speaker 2

Mark Riedl is an Assistant Professor in the Georgia Tech School of Interactive Computing and director of the Entertainment Intelligence Lab. Dr. Riedl's research focuses on the intersection of artificial intelligence, virtual worlds, and storytelling. The principle research question Dr. Riedl addresses through his research is: how can intelligent computational systems reason about and autonomously create engaging experiences for users of virtual worlds and computer games. Dr. Riedl earned a PhD degree in 2004 from North Carolina State University, where he developed intelligent systems for generating stories and managing interactive user experiences in computer games. From 2004 to 2007, Dr. Riedl was a Research Scientist at the University of Southern California Institute for Creative Technologies where he researched and developed interactive, narrative-based training systems. Dr. Riedl joined the Georgia Tech College of Computing in 2007 and in 2011 he received a DARPA Young Faculty Award for his work on artificial intelligence, narrative, and virtual worlds. His research is supported by the NSF, DARPA, the U.S. Army, and Disney.

Intelligent Narrative Generation: From Cognition to Crowdsourcing

Abstract

Storytelling is a pervasive part of the human experience–we as humans tell stories to communicate, inform, entertain, and educate. But what about computational systems? There are many applications for which we would also like intelligent system to reason about, understand, and create narrative structures: from recognition to question-answering, from entertainment to education. In this talk I will look at one particular aspect of computational modeling of narrative: automated story generation, the problem of creating novel narrative fabula event sequences for dramatic, pedagogical, or other purposes. I will trace the evolution of fabula story generation from its roots in cognitive systems to data-driven techniques and crowdsourcing. I will speculate on how systems may eventually learn how to create and tell stories from interacting with humans and literature.