

NAACL 2013

**Proceedings of the
Workshop on Language Analysis in Social Media**

13 June 2013
Atlanta, Georgia

©2013 The Association for Computational Linguistics

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

Introduction

These proceedings contain the papers presented at the workshop on Language Analysis in Social Media (LASM 2013). The workshop was held in Atlanta, Georgia, USA and hosted in conjunction with the 2013 Conference of the North American Chapter of the Association for Computational Linguistics-Human Language Technologies (NAACL-HLT 2013).

Over the last few years, there has been a growing public and enterprise interest in social media and their role in modern society. At the heart of this interest is the ability for users to create and share content via a variety of platforms such as blogs, micro-blogs, collaborative wikis, multimedia sharing sites, and social networking sites. The unprecedented volume and variety of user-generated content as well as the user interaction network constitute new opportunities for understanding social behavior and building socially-aware systems.

The Workshop Committee received several submissions for LASM 2013 from around the world. Each submission was reviewed by up to four reviewers. For the final workshop program, and for inclusion in these proceedings, nine regular papers, of 11 pages each, were selected.

This workshop was intended to serve as a forum for sharing research efforts and results in the analysis of language with implications for fields such as computational linguistics, sociolinguistics and psycholinguistics. We invited original and unpublished research papers on all topics related the analysis of language in social media, including the following topics:

- What are people talking about on social media?
- How are they expressing themselves?
- Why do they scribe?
- Natural language processing techniques for social media analysis
- Language and network structure: How do language and social network properties interact?
- Semantic Web / Ontologies / Domain models to aid in social data understanding
- Language across verticals
- Characterizing Participants via Linguistic Analysis
- Language, Social Media and Human Behavior

This workshop would not have been possible without the hard work of many people. We would like to thank all Program Committee members and external reviewers for their effort in providing high-quality reviews in a timely manner. We thank all the authors who submitted their papers, as well as the authors whose papers were selected, for their help with preparing the final copy. Many thanks to our industrial partners.

We are in debt to NAACL-HLT 2013 Workshop Chairs Luke Zettlemoyer and Sujith Ravi. We would also like to thank our industry partners Microsoft Research, IBM Almaden and NLP Technologies.

May 2013

Atefeh Farzindar

Michael Gamon

Meena Nagarajan

Diana Inkpen

Cristian Danescu-Niculescu-Mizil

Organizers:

Cristian Danescu-Niculescu-Mizil, Stanford University and Max Planck Institute SWS
Atefeh Farzindar, NLP Technologies
Michael Gamon, Microsoft Research
Diana Inkpen, University of Ottawa
Meenakshi Nagarajan, IBM Almaden

Program Committee:

Cindy Chung (University of Texas)
Munmun De Choudhury (Microsoft Research)
Jennifer Foster (Dublin City University)
Daniel Gruhl (IBM Research)
Kevin Haas (Microsoft)
Guy Lapalme (Université de Montréal)
Eduarda Mendes Rodrigues (University of Porto)
Alena Neviarouskaya (University of Tokyo)
Nicolas Nicolov (Microsoft)
Alexander Osherenko (University of Augsburg)
Patrick Pantel (Microsoft Research)
Alan Ritter (University of Washington)
Mathieu Roche (Université de Montpellier)
Victoria Rubin (University of Western Ontario)
Hassan Sayyadi (University of Maryland)
Amit Sheth (Wright State)
Scott Spangler (IBM Research)
Mike Thelwall (University of Wolverhampton)
Alessandro Valitutti (University of Helsinki)
Julien Velcin (Université de Lyon)
Emre Kiciman (Microsoft Research)
Valerie Shalin (Wright State)
Ian Soboroff (NIST)

Invited Speaker:

Mor Naaman, Rutgers University

Table of Contents

<i>Does Size Matter? Text and Grammar Revision for Parsing Social Media Data</i> Mohammad Khan, Markus Dickinson and Sandra Kuebler	1
<i>Phonological Factors in Social Media Writing</i> Jacob Eisenstein	11
<i>A Preliminary Study of Tweet Summarization using Information Extraction</i> Wei Xu, Ralph Grishman, Adam Meyers and Alan Ritter	20
<i>Really? Well. Apparently Bootstrapping Improves the Performance of Sarcasm and Nastiness Classifiers for Online Dialogue</i> Stephanie Lukin and Marilyn Walker	30
<i>Topical Positioning: A New Method for Predicting Opinion Changes in Conversation</i> Ching-Sheng Lin, Samira Shaikh, Jennifer Stromer-Galley, Jennifer Crowley, Tomek Strzalkowski and Veena Ravishankar	41
<i>Sentiment Analysis of Political Tweets: Towards an Accurate Classifier</i> Akshat Bakliwal, Jennifer Foster, Jennifer van der Puil, Ron O'Brien, Lamia Tounsi and Mark Hughes	49
<i>A Case Study of Sockpuppet Detection in Wikipedia</i> Thamar Solorio, Ragib Hasan and Mainul Mizan	59
<i>Towards the Detection of Reliable Food-Health Relationships</i> Michael Wiegand and Dietrich Klakow	69
<i>Translating Government Agencies' Tweet Feeds: Specificities, Problems and (a few) Solutions</i> Fabrizio Gotti, Philippe Langlais and Atefeh Farzindar	80

Conference Program

Thursday, June 13, 2013

- 9:00–9:15 Introductions
- 9:15–10:30 Invited Key Note, Prof. Mor Naaman
- 10:30–11:00 Coffee Break
- 11:00–11:30 *Does Size Matter? Text and Grammar Revision for Parsing Social Media Data*
Mohammad Khan, Markus Dickinson and Sandra Kuebler
- 11:30–12:00 *Phonological Factors in Social Media Writing*
Jacob Eisenstein
- 12:00–12:30 *A Preliminary Study of Tweet Summarization using Information Extraction*
Wei Xu, Ralph Grishman, Adam Meyers and Alan Ritter
- 12:30–2:00 Lunch
- 2:00–2:30 *Really? Well. Apparently Bootstrapping Improves the Performance of Sarcasm and Nastiness Classifiers for Online Dialogue*
Stephanie Lukin and Marilyn Walker
- 2:30–3:00 *Topical Positioning: A New Method for Predicting Opinion Changes in Conversation*
Ching-Sheng Lin, Samira Shaikh, Jennifer Stromer-Galley, Jennifer Crowley, Tomek Strzalkowski and Veena Ravishankar
- 3:00–3:30 *Sentiment Analysis of Political Tweets: Towards an Accurate Classifier*
Akshat Bakliwal, Jennifer Foster, Jennifer van der Puil, Ron O'Brien, Lamia Tounsi and Mark Hughes
- 3:30–3:45 Coffee Break
- 3:45–4:15 *A Case Study of Sockpuppet Detection in Wikipedia*
Tamar Solorio, Ragib Hasan and Mainul Mizan
- 4:15–4:45 *Towards the Detection of Reliable Food-Health Relationships*
Michael Wiegand and Dietrich Klakow
- 4:45–5:15 *Translating Government Agencies' Tweet Feeds: Specificities, Problems and (a few) Solutions*
Fabrizio Gotti, Philippe Langlais and Atefeh Farzindar
- 5:15 Closing Remarks