

ACL HLT 2011

**Workshop on Language in Social Media
LSM 2011**

Proceedings of the Workshop

23 June, 2011
Portland, Oregon, USA

Production and Manufacturing by
Omnipress, Inc.
2600 Anderson Street
Madison, WI 53704 USA

©2011 The Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-932432-96-1

Introduction

Welcome to the ACL Workshop on Language in Social Media (LSM 2011)!

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, social networking sites etc. The volume and variety of user-generated content (UGC) and the user participation network behind it are creating new opportunities for understanding web-based practices and building socially intelligent and personalized applications. Investigations around social data can be broadly categorized along the following dimensions:

- (a) understanding aspects of the user-generated content
- (b) modeling and observing the user network that the content is generated in and
- (c) characterizing individuals and groups that produce and consume the content.

The goals for this workshop are to focus on sharing research efforts and results in the area of understanding language usage on social media.

While there is a rich body of previous work in processing textual content, certain characteristics of UGC on social media introduce challenges in their analyses. A large portion of language found in UGC is in the Informal English domain — a blend of abbreviations, slang and context specific terms; lacking in sufficient context and regularities and delivered with an indifferent approach to grammar and spelling. Traditional content analysis techniques developed for a more formal genre like news, Wikipedia or scientific articles do not necessarily translate well to UGC. Consequently, well-understood problems such as information extraction, search or monetization on the Web are facing pertinent challenges and need to be revisited.

Meena Nagarajan and Michael Gamon

Organizers:

Meena Nagarajan (IBM Research)
Michael Gamon (Microsoft Research)

Program Committee:

John Breslin (University of Galway)
Cindy Chung (University of Texas)
Munmun De Choudhury (Arizona State University)
Cristian Danescu-Niculescu-Mizil (Cornell University)
Susan Dumais (Microsoft Research)
Jennifer Foster (Dublin City University)
Sam Gosling (University of Texas)
Julia Grace (IBM Research)
Daniel Gruhl (IBM Research)
Kevin Haas (Microsoft)
Emre Kiciman (Microsoft Research)
Nicolas Nicolov (Microsoft)
Daniel Ramage (Stanford University)
Alan Ritter (University of Washington)
Christine Robson (IBM Research)
Hassan Sayyadi (University of Maryland)
Valerie Shalin (Wright State University)
Amit Sheth (Wright State University)
Ian Soboroff (NIST)
Hari Sundaram (Arizona State University)
Scott Spangler (IBM Research)
Smaranda Muresan (Rutgers University)

Additional Reviewers:

Kristina Toutanova (Microsoft Research)
Josephine Griffith (NUI Galway)
Hemant Purohit (Wright State University)
Lu Chen (Wright State University)
Ashutosh Jadhav (Wright State University)

Invited Speaker:

Susan C. Herring (Indiana University)

Table of Contents

<i>Automating Analysis of Social Media Communication: Insights from CMDA</i> Susan Herring	1
<i>How can you say such things?!?: Recognizing Disagreement in Informal Political Argument</i> Rob Abbott, Marilyn Walker, Pranav Anand, Jean E. Fox Tree, Robeson Bowmani and Joseph King	2
<i>What pushes their buttons? Predicting comment polarity from the content of political blog posts</i> Ramnath Balasubramanian, William W. Cohen, Doug Pierce and David P. Redlawsk	12
<i>Contextual Bearing on Linguistic Variation in Social Media</i> Stephan Gouws, Donald Metzler, Congxing Cai and Eduard Hovy	20
<i>Sentiment Analysis of Twitter Data</i> Apoorv Agarwal, Boyi Xie, Ilia Vovsha, Owen Rambow and Rebecca Passonneau	30
<i>Detecting Forum Authority Claims in Online Discussions</i> Alex Marin, Bin Zhang and Mari Ostendorf	39
<i>Annotating Social Acts: Authority Claims and Alignment Moves in Wikipedia Talk Pages</i> Emily M. Bender, Jonathan T. Morgan, Meghan Oxley, Mark Zachry, Brian Hutchinson, Alex Marin, Bin Zhang and Mari Ostendorf	48
<i>Analyzing the Dynamic Evolution of Hashtags on Twitter: a Language-Based Approach</i> Evandro Cunha, Gabriel Magno, Giovanni Comarela, Virgilio Almeida, Marcos André Gonçalves and Fabricio Benevenuto	58
<i>Why is "SXSW" trending? Exploring Multiple Text Sources for Twitter Topic Summarization</i> Fei Liu, Yang Liu and Fuliang Weng	66
<i>Language use as a reflection of socialization in online communities</i> Dong Nguyen and Carolyn P. Rosé	76
<i>Email Formality in the Workplace: A Case Study on the Enron Corpus</i> Kelly Peterson, Matt Hohensee and Fei Xia	86

Conference Program

9:00 **Opening remarks**

Keynote

9:15 *Automating Analysis of Social Media Communication: Insights from CMDA*
Susan Herring

10:15 **Coffee break**

Session 1

10:30 *How can you say such things?!?: Recognizing Disagreement in Informal Political Argument*
Rob Abbott, Marilyn Walker, Pranav Anand, Jean E. Fox Tree, Robeson Bowmani and Joseph King

11:00 *What pushes their buttons? Predicting comment polarity from the content of political blog posts*
Ramnath Balasubramanyan, William W. Cohen, Doug Pierce and David P. Redlawsk

11:30 *Contextual Bearing on Linguistic Variation in Social Media*
Stephan Gouws, Donald Metzler, Congxing Cai and Eduard Hovy

12:00 *Sentiment Analysis of Twitter Data*
Apoorv Agarwal, Boyi Xie, Ilia Vovsha, Owen Rambow and Rebecca Passonneau

12:30 **Lunch break**

Session 2

1:30 *Detecting Forum Authority Claims in Online Discussions*
Alex Marin, Bin Zhang and Mari Ostendorf

2:00 *Annotating Social Acts: Authority Claims and Alignment Moves in Wikipedia Talk Pages*
Emily M. Bender, Jonathan T. Morgan, Meghan Oxley, Mark Zachry, Brian Hutchinson,
Alex Marin, Bin Zhang and Mari Ostendorf

2:30 *Analyzing the Dynamic Evolution of Hashtags on Twitter: a Language-Based Approach*
Evandro Cunha, Gabriel Magno, Giovanni Comarela, Virgilio Almeida, Marcos André
Gonçalves and Fabricio Benevenuto

3:00 *Why is "SXSW" trending? Exploring Multiple Text Sources for Twitter Topic Summariza-
tion*
Fei Liu, Yang Liu and Fuliang Weng

Session 3

4:00 *Language use as a reflection of socialization in online communities*
Dong Nguyen and Carolyn P. Rosé

4:30 *Email Formality in the Workplace: A Case Study on the Enron Corpus*
Kelly Peterson, Matt Hohensee and Fei Xia

5:00 **Discussion and Closing Remarks**