

SocialNLP 2015@NAACL

**The Third International Workshop on
Natural Language Processing for Social Media**

Proceedings of the Workshop

June 5, 2015
Denver, Colorado, USA

©2015 The Association for Computational Linguistics

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 978-1-941643-48-8

SocialNLP 2015@NAACL Chairs' Welcome

It is our great pleasure to welcome you to the Third Workshop on Natural Language Processing for Social Media – SocialNLP'15, associated with NAACL 2015. SocialNLP is a new inter-disciplinary area of natural language processing (NLP) and social computing. There are three plausible directions of SocialNLP: (1) addressing issues in social computing using NLP techniques; (2) solving NLP problems using information from social media; and (3) handling new problems related to both social computing and natural language processing.

Through this workshop, we anticipate to provide a platform for research outcome presentation and head-to-head discussion in the area of SocialNLP, with the hope to combine the insight and experience of prominent researchers from both NLP and social computing domains to contribute to the area of SocialNLP jointly. Also, selected and expanded versions of papers presented at SocialNLP will be published in two follow-on Special Issues of Springer Cognitive Computation (CogComp) and the International Journal of Computational Linguistics and Chinese Language Processing (IJCLCLP).

The submissions to this year's workshop were again of high quality and we had a competitive selection process. We received 10 submissions from Asia, Europe, and the United States., and due to a rigorous review process, we only accepted 5 of them. Thus the acceptance rate was 50 percent. The workshop papers cover a broad range of SocialNLP-related topics, such as location name disambiguation, microblog and mobile game text processing, product mining, and social media user analysis. We had a total of 27 program committee members, and each submission is evaluated by at least 3 PC members. We warmly thank our PC members for the timely reviews and constructive comments.

We are delighted to have two keynote speeches this year. Prof. Jacob Eisenstein, from Georgia Institute of Technology, will give a talk entitled "Variation and Change in Social Media Language"; Prof. Michael C. Frank, from Stanford University, will give a talk entitled "Predicting Pragmatic Reasoning about Language Use in Context". We also encourage attendees to attend the keynote talk presentations. These valuable and insightful talks can and will guide us to a better understanding of the future.

Putting together SocialNLP 2015 was a team effort. We first thank the authors for providing the content of the program. We are grateful to the program committee members, who worked very hard in reviewing papers and providing feedback for authors. Finally, we especially thank the Workshop Committee Chairs Prof. Matt Post and Prof. Adam Lopez.

We hope you keep supporting SocialNLP workshop and enjoying it!

Organizers of SocialNLP 2015,

Shou-De Lin, Lun-Wei Ku, Cheng-Te Li and Erik Cambria

Organizers:

Shou-de Lin (National Taiwan University, Taiwan)
Lun-Wei Ku (Academia Sinica, Taiwan)
Cheng-Te Li (Academia Sinica, Taiwan)
Erik Cambria (Nanyang Technological University, Singapore)

Program Committee:

Alexandra Balahur (European Commission's Joint Research Centre)
Chia-Hui Chang (National Central University, Taiwan)
Berlin Chen (National Taiwan Normal University, Taiwan)
Hsin-Hsi Chen (National Taiwan University, Taiwan)
Munmun De Choudhury (Georgia Institute of Technology, USA)
Min-Yuh Day (Tamkang University, Taiwan)
Amitava Das (University of North Texas, USA)
Dipankar Das (Jadavpur University, India)
Jennifer Foster (Dublin City University, Ireland)
June-Jei Kuo (National Chung Hsing University, Taiwan)
Chuan-Jie Lin (National Taiwan Ocean University, Taiwan)
Rafal Rzepka (Hokkaido University, Japan)
Yohei Seki (University of Tsukuba, Japan)
Marina Sokolova (University of Ottawa, USA)
Daniela Stockmann (Leiden University, The Netherlands)
Keh-Yih Su (Academia Sinica, Taiwan)
Ming-Feng Tsai (National ChengChi University, Taiwan)
Chi Wang (Microsoft Research, USA)
Hsin-Min Wang (Academia Sinica, Taiwan)
Jenq-Haur Wang (National Taipei University of Technology, Taiwan)
Shih-Hung Wu (Chaoyang University of Technology, Taiwan)
Yejun Wu (Louisiana State University, USA)
Yunqing Xia (Tsinghua University, China)
Ruifeng Xu (Harbin Institute of Technology, China)
Rui Yan (Baidu Inc., China)
Yi-Hsuan Yang (Academia Sinica, Taiwan)
Kevin Zhang (Beijing Institute of Technology, China)

Keynote Speaker:

Jacob Eisenstein (Georgia Institute of Technology, USA)
Michael C. Frank (Stanford University, USA)

Supported by Asia Federation of Natural Language Processing

Keynote Speech (I)

Keynote Speaker (Morning Session):

Jacob Eisenstein (Georgia Institute of Technology, USA)

Title:

Variation and Change in Social Media Language

Abstract:

Social media is sometimes described as a new domain, genre, or task for natural language processing. This suggests that it has specific properties that distinguish it from other sources of text. I will argue that there are exactly two such properties: variation and change. NLP research has historically focused on genres such as newstext, where there is strong pressure towards standardization. Far less pressure exists in social media, and so we must contend with variation on all levels of the linguistic spectrum. This variation enables authors to mark a diverse array of social relationships and identities, and with this increasingly important interpersonal role, online writing becomes enmeshed in complex social processes that lead to instability and change. The inherently dynamic nature of social media language is why we can no longer annotate our way to high accuracy NLP, so learning from unlabeled data will be increasingly critical. Finally, while variation and change pose challenges, they also offer new opportunities for deepening our understanding of both language and social processes. I will describe our recent work on mining four years of Twitter data to uncover macro-scale pathways of linguistic influence among American cities.

Speaker Biography:

Jacob Eisenstein is an Assistant Professor in the School of Interactive Computing at Georgia Tech. He works on statistical natural language processing, focusing on computational sociolinguistics, social media analysis, discourse, and machine learning. He is a recipient of the NSF CAREER Award, a member of the Air Force Office of Scientific Research (AFOSR) Young Investigator Program, and was a SICSA Distinguished Visiting Fellow at the University of Edinburgh. His work has also been supported by the National Institutes for Health, the National Endowment for the Humanities, and Google. Jacob was a Postdoctoral researcher at Carnegie Mellon and the University of Illinois. He completed his Ph.D. at MIT in 2008, winning the George M. Sprowls dissertation award. Jacob's research has been featured in the New York Times, National Public Radio, and the BBC. Thanks to his brief appearance in *If These Knishes Could Talk*, Jacob has a Bacon number of 2.

Keynote Speech (II)

Keynote Speaker (Afternoon Session):

Michael C. Frank (Stanford University, USA)

Title:

Predicting Pragmatic Reasoning about Language Use in Context

Abstract:

A short, ambiguous message can convey a lot of information, provided the listener is willing to make inferences based on assumptions about the speaker and the context of the message. These sorts of pragmatic inferences are critical in facilitating efficient human communication, and have been characterized informally using tools like Grice's conversational maxims. In this talk, I'll describe our work on a new, probabilistic framework for referential communication in context. This framework shows good fit to adults' and children's judgments across many experiments, provides extensions to a variety of complex linguistic phenomena, and resolves some important puzzles about language processing. I'll end by describing how we have begun to test this framework using data from large-scale corpora of social media conversations.

Speaker Biography:

Michael C. Frank is Associate Professor of Psychology at Stanford University. He earned his BS from Stanford University in Symbolic Systems in 2005 and his PhD from MIT in Brain and Cognitive Sciences in 2010. He studies both adults' language use and children's language learning and how both of these interact with social cognition. His work uses behavioral experiments, computational tools, and novel measurement methods including large-scale web-based studies, eye-tracking, and head-mounted cameras.

Table of Contents

<i>Location Name Disambiguation Exploiting Spatial Proximity and Temporal Consistency</i> Takashi Awamura, Daisuke Kawahara, Eiji ARAMAKI, Tomohide Shibata and Sadao Kurohashi 1	
<i>Paraphrase Identification and Semantic Similarity in Twitter with Simple Features</i> Ngoc Phuoc An Vo, Simone Magnolini and Octavian Popescu	10
<i>A Language Detection System for Short Chats in Mobile Games</i> Pidong Wang, Nikhil Bojja and Shivasankari Kannan	20
<i>Long Nights, Rainy Days, and Misspent Youth: Automatically Extracting and Categorizing Occasions Associated with Consumer Products.</i> David Bracewell	29
<i>A Deep Learning and Knowledge Transfer Based Architecture for Social Media User Characteristic Determination</i> Matthew Riemer, Sophia Krasikov and Harini Srinivasan	39

Conference Program

Friday, June 5, 2015

09:20–09:30 *Opening*

09:30–10:30 *Keynote Speech (I)*

09:30–10:30 *Variation and Change in Social Media Language*
Jacob Eisenstein (Georgia Institute of Technology)

10:30–11:00 *Coffee Break (I)*

11:00–12:00 **Natural Language Processing Session**

11:00–11:30 *Location Name Disambiguation Exploiting Spatial Proximity and Temporal Consistency*

Takashi Awamura, Daisuke Kawahara, Eiji ARAMAKI, Tomohide Shibata and Sadao Kurohashi

11:30–12:00 *Paraphrase Identification and Semantic Similarity in Twitter with Simple Features*
Ngoc Phuoc An Vo, Simone Magnolini and Octavian Popescu

12:00–13:30 *Lunch*

13:30–14:00 *Panel Discussion*

14:00–15:00 *Keynote Speech (II)*

14:00–15:00 *Predicting Pragmatic Reasoning about Language Use in Context*
Michael C. Frank (Stanford University)

15:00–15:30 *Coffee Break (II)*

Friday, June 5, 2015 (continued)

15:30–17:00 Social Media Session

15:30–16:00 *A Language Detection System for Short Chats in Mobile Games*

Pidong Wang, Nikhil Bojja and Shivasankari Kannan

16:00–16:30 *Long Nights, Rainy Days, and Misspent Youth: Automatically Extracting and Categorizing Occasions Associated with Consumer Products.*

David Bracewell

16:30–17:00 *A Deep Learning and Knowledge Transfer Based Architecture for Social Media User Characteristic Determination*

Matthew Riemer, Sophia Krasikov and Harini Srinivasan

17:00–17:20 Best Paper Award/Closing