

EMNLP 2016

**Conference on
Empirical Methods in
Natural Language Processing**

**Proceedings of The Fourth International Workshop
on Natural Language Processing for Social Media**

November 1, 2016
Austin, Texas, USA

©2016 The Association for Computational Linguistics

Order copies of this and other ACL proceedings 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 978-1-945626-32-6

SocialNLP@EMNLP2016 Chairs' Welcome

It is our great pleasure to welcome you to the Fourth ACM Workshop on Natural Language Processing for SocialMedia – SocialNLP'16, associated with EMNLP 2016. SocialNLP is an inter-disciplinary area of natural language processing (NLP) and social computing. We hold SocialNLP twice a year: one in the NLP venue, the other in the associated venue such as those for web technology or artificial intelligence. 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. The submissions to this year's workshop were again of high quality and we had a competitive selection process. We received 32 submissions from Asia, Europe, and the United States, and due to a rigorous review process, we only accepted 6 as long oral papers and 7 as short oral papers. Thus the acceptance rate was 40 percent. We are delighted to have Prof. Cristian Danescu-Niculescu-Mizil, from Cornell University, as our keynote speaker. We also encourage attendees to attend the keynote talk presentation. The valuable and insightful talk can and will guide us to a better understanding of the future. Putting together SocialNLP 2016 was a team effort. We first thank the authors for providing the quality 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. Annie Louis and Prof. Greg Kondrak.

We hope you join our community and enjoy the workshop!

Organizers

Jane Yung-jen Hsu, National Taiwan University, Taiwan

Lun-Wei Ku, Academia Sincia, Taiwan

Cheng-Te Li, National Cheng Kung University, Taiwan

Organizers:

Jane Yung-jen Hsu, National Taiwan University, Taiwan
Lun-Wei Ku, Academia Sinica, Taiwan
Cheng-Te Li, National Cheng-Kung University, Taiwan

Program Committee:

Zeljko Agic, University of Copenhagen
Nikolaos Aletras, Amazon UK
Tim Althoff, Stanford University
Hadi Amiri, University of Maryland
Ion Androutsopoulos, Athens University of Economics and Business
Alexandra Balahur, European Commission Joint Research Centre
Roberto Basili, University of Rome Tor Vergata
Fabrício Benevenuto, Federal University of Minas Gerais
Kalina Bontcheva, University of Sheffield
Taylor Cassidy, US Army Research Laboratory
Berlin Chen, National Taiwan Normal University
Hsin-Hsi Chen, National Taiwan University
John Chen, Interactions LLC
Hai Leong Chieu, DSO National Laboratories
Monojit Choudhury, Microsoft Research, India
Lei Cui, Microsoft Research
Aron Culotta, Illinois Institute of Technology
Pradipto Das, Rakuten Institute of Technology
Leon Derczynski, The University of Sheffield
Marco Dinarelli, Lattice-CNRS
Koji Eguchi, Kobe University
Michael Elhadad, Ben-Gurion University of the Negev
Hugo Jair Escalante, INAOE
Wei Gao, Qatar Computing Research Institute
Spandana Gella, University of Edinburgh
Alastair Gill, King's College London
Weiwei Guo, Yahoo! Labs
Scott Hale, University of Oxford
William Hamilton, Stanford University
Bo Han, IBM Research
Catherine Havasi, Luminoso, MIT
Yulan He, Aston University
Michael Heilman, Civis Analytics
Graeme Hirst, University of Toronto
John Henderson, MITRE

Tuan-Anh Hoang, Singapore Management University
Wen-Lian Hsu, Academia Sinica
Ruihong Huang, Texas A&M University
Ting-Hao Huang, Carnegie Mellon University
Iustina Ilisei, Cognizant Technology Solutions Corp.
Yangfeng Ji, Georgia Tech
Jing Jiang, Singapore Management University
Anders Johannsen, University of Copenhagen
David Jurgens, Stanford University
Nobuhiro Kaji, Yahoo! Japan Corp.
Pallika Kanani, Oracle Labs
Emre Kiciman, Microsoft Research
Dongwoo Kim, ANU
Suin Kim, Korea Advanced Institute of Science and Technology
Roman Klinger, University of Stuttgart
Lingpeng Kong, Carnegie Mellon University
June-Jei Kuo, National Chung Hsing University
Tsun-Ting Kuo, University of California, San Diego
Patrik Lambert, Universitat Pompeu Fabra
Man Lan, East China Normal University
Kyumin Lee, Utah State University
Sungjin Lee, Yahoo! Labs
Haibo Li, Nuance Communications
Shou-De Lin, National Taiwan University
Yu-Ru Lin, University of Pittsburgh
Chuan-Jie Lin, National Taiwan Ocean University
Kang Liu, Chinese Academy of Sciences
Zhiyuan Liu, Tsinghua University
Bin Lu, Google Inc.
Zhunchen Luo, China Defense Science and Technology Information Center
Bruno Martins, University of Lisbon
Diana Maynard, University of Sheffield
Karo Moilanen, University of Oxford
Manuel Montes-y-Gómez, National Institute of Astrophysics, Optics and Electronics
Edward Newell, McGill University
Dong Nguyen, University of Twente
Scott Nowson, Xerox Research Centre Europe
Miles Osborne, Bloomberg
George Paliouras, NCSR, Demokritos
Harris Papageorgiou, ATHENA RC
Michael Paul, University of Colorado Boulder
Barbara Plank, University of Copenhagen
Stephan Raaijmakers, TNO, The Netherlands
Sravana Reddy, Wellesley College
Saurav Sahay, Intel Labs
Hassan Saif, The Open University

Yohei Seki, University of Tsukuba
Mário J. Silva, Universidade de Lisboa
Yanchuan Sim, Carnegie Mellon University
Jan Snajder, University of Zagreb
Veselin Stoyanov, Facebook Inc.
Carlo Strapparava, FBK-irst
Keh-Yih Su, Academia Sinica
Hiroya Takamura, Tokyo Institute of Technology
Xavier Tannier, Université Paris-Sud, Université Paris-Saclay, LIMSIS, CNRS
Ming-Feng Tsai, National Chengchi University
Paola Velardi, University of Roma La Sapienza
Marc Verhagen, Brandeis University
Svitlana Volkova, PNNL
Xiaojun Wan, Peking University
Hsin-Min Wang, Academia Sinica
Jenq-Haur Wang, National Taipei University of Technology
William Yang Wang, Carnegie Mellon University
Ingmar Weber, Qatar Computing Research Institute
Albert Weichselbraun, University of Applied Sciences Chur
Robert West, Stanford University
Janyce Wiebe, University of Pittsburgh
Ruifeng Xu, Harbin Institute of Technology
Yi Yang, Georgia Tech
Yi-Hsuan Yang, Academia Sinica
Bei Yu, Syracuse University
Liang-Chih Yu, Yuan Ze University
Nicholas Jing Yuan, Microsoft Research
Zhe Zhang, IBM Watson
Hua-Ping Zhang, Beijing Institute of Technology
Xin Zhao, Renmin University of China
Deyu Zhou, Southeast University
Jun Zhu, Tsinghua University

Invited Speaker:

Cristian Danescu-Niculescu-Mizil, Cornell University

Table of Contents

<i>Identifying and Categorizing Disaster-Related Tweets</i>	
Kevin Stowe, Michael J. Paul, Martha Palmer, Leysia Palen and Kenneth Anderson	1
<i>Identifying Eyewitness News-worthy Events on Twitter</i>	
Erika Doggett and Alejandro Cantarero	7
<i>Why Do They Leave: Modeling Participation in Online Depression Forums</i>	
Farig Sadeque, Ted Pedersen, Thamar Solorio, Prasha Shrestha, Nicolas Rey-Villamizar and Steven Bethard	14
<i>Twitter at the Grammys: A Social Media Corpus for Entity Linking and Disambiguation</i>	
Mark Dredze, Nicholas Andrews and Jay DeYoung	20
<i>Steps Toward Automatic Understanding of the Function of Affective Language in Support Groups</i>	
Amit Navindgi, Caroline Brun, Cécile Boulard Masson and Scott Nowson	26
<i>Detecting Social Roles in Twitter</i>	
Sunghwan Mac Kim, Stephen Wan and Cecile Paris	34
<i>Identifying Sensible Participants in Online Discussions</i>	
Siddharth Jain	41
<i>emoji2vec: Learning Emoji Representations from their Description</i>	
Ben Eisner, Tim Rocktäschel, Isabelle Augenstein, Matko Bosnjak and Sebastian Riedel	48
<i>Learning Latent Local Conversation Modes for Predicting Comment Endorsement in Online Discussions</i>	
Hao Fang, Hao Cheng and Mari Ostendorf	55
<i>Witness Identification in Twitter</i>	
Rui Fang, Armineh Nourbakhsh, XIAOMO LIU, Sameena Shah and Quanzhi Li	65
<i>How Do I Look? Publicity Mining From Distributed Keyword Representation of Socially Infused News Articles</i>	
Yu-Lun Hsieh, Yung-Chun Chang, Chun-Han Chu and Wen-Lian Hsu	74
<i>Hierarchical Character-Word Models for Language Identification</i>	
Aaron Jaech, George Mulcaire, Shobhit Hathi, Mari Ostendorf and Noah A. Smith	84
<i>Human versus Machine Attention in Document Classification: A Dataset with Crowdsourced Annotations</i>	
Nikolaos Pappas and Andrei Popescu-Belis	94

Workshop Program

Tuesday, November 1, 2016

08:50–09:00 **Opening**

09:00–10:00 **Keynote Speech**

09:00–10:00 *Social Cues in Conversational Dynamics*
Cristian Danescu-Niculescu-Mizil

10:00–10:30 **Short Paper Session I**

10:00–10:15 *Identifying and Categorizing Disaster-Related Tweets*
Kevin Stowe, Michael J. Paul, Martha Palmer, Leysia Palen and Kenneth Anderson

10:15–10:30 *Identifying Eyewitness News-worthy Events on Twitter*
Erika Doggett and Alejandro Cantarero

10:30–11:00 *Coffee Break*

11:00–12:15 **Short Paper Session II**

11:00–11:15 *Why Do They Leave: Modeling Participation in Online Depression Forums*
Farig Sadeque, Ted Pedersen, Thamar Solorio, Prasha Shrestha, Nicolas Rey-Villamizar and Steven Bethard

11:15–11:30 *Twitter at the Grammys: A Social Media Corpus for Entity Linking and Disambiguation*
Mark Dredze, Nicholas Andrews and Jay DeYoung

11:30–11:45 *Steps Toward Automatic Understanding of the Function of Affective Language in Support Groups*
Amit Navindgi, Caroline Brun, Cécile Boulard Masson and Scott Nowson

11:45–12:00 *Detecting Social Roles in Twitter*
Sunghwan Mac Kim, Stephen Wan and Cecile Paris

12:00–12:15 *Identifying Sensible Participants in Online Discussions*
Siddharth Jain

Tuesday, November 1, 2016 (continued)

12:30–14:00 *Lunch*

14:00–15:30 **Long Paper Session I**

14:00–14:30 *emoji2vec: Learning Emoji Representations from their Description*
Ben Eisner, Tim Rocktäschel, Isabelle Augenstein, Matko Bosnjak and Sebastian Riedel

14:30–15:00 *Learning Latent Local Conversation Modes for Predicting Comment Endorsement in Online Discussions*
Hao Fang, Hao Cheng and Mari Ostendorf

15:00–15:30 *Witness Identification in Twitter*
Rui Fang, Armineh Nourbakhsh, XIAOMO LIU, Sameena Shah and Quanzhi Li

15:30–16:00 *Coffee Break / Poster Session and Discussion*

16:00–17:30 **Long Paper Session II**

16:00–16:30 *How Do I Look? Publicity Mining From Distributed Keyword Representation of Socially Infused News Articles*
Yu-Lun Hsieh, Yung-Chun Chang, Chun-Han Chu and Wen-Lian Hsu

16:30–17:00 *Hierarchical Character-Word Models for Language Identification*
Aaron Jaech, George Mulcaire, Shobhit Hathi, Mari Ostendorf and Noah A. Smith

17:00–17:30 *Human versus Machine Attention in Document Classification: A Dataset with Crowdsourced Annotations*
Nikolaos Pappas and Andrei Popescu-Belis

Tuesday, November 1, 2016 (continued)

17:30–17:40 Award and Closing