

SocialNLP 2021

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

**Proceedings of the Workshop**

June 10, 2021

©2021 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](mailto:acl@aclweb.org)

ISBN 978-1-954085-32-9

## SocialNLP 2021@NAACL Chairs' Welcome

It is our great pleasure to welcome you to the Ninth Workshop on Natural Language Processing for Social Media-SocialNLP 2021, associated with NAACL 2021. 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. This year the other version has been successfully held in conjunction with TheWebConf 2021 (formerly WWW), and we are very happily looking forward the NLP version in NAACL 2021. We are very glad that the number of submissions to this year's workshop keeps increasing this year, and the submissions themselves were still of high quality with the accepted threshold 3.33 (maximum 5), which again leads to a competitive selection process. We received submissions from Asia, Europe, and the United States. Considering the review process is rigorous and we want to encourage authors to participate the online workshop, we accepted 16 oral papers and thus the acceptance rate was 55 percent. These exciting papers include novel and practical topics for researchers working on NLP for social media, such as bias mitigation, domain transfer, and dataset constructed for the newly emerged research problems. We believe they will benefit our research community.

Besides the main workshop, we are having this year a new EmotionX challenge, Fake-EmoReact. At the time we compose this proceedings, we already have 20 international teams registered and the challenge is still ongoing. We have a special session for this challenge to exchange related ideas and experience in the workshop. We hope this challenge series can bring participants from the research problem to the real solution.

This year we are excited to have Prof. Dan Goldwasser from Purdue University and Prof. Tim Weninger from University of Notre Dame as our keynote speakers. As one of our beautiful tradition, all the authors provide their poster in gather.town. We encourage attendees to (virtually) attend our keynote speech, oral sessions as well as the poster session to have more discussions with outstanding researchers.

Putting together SocialNLP 2021 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 to authors. For a lot of tedious work coming from the challenge, we thank the challenge co-chairs Prof Hon-Han Shuai and Dr. Ming Sun for their great effort. We also thank Mr. Chien-Kun Huang and Mr. Yi-Ting Chang for their assistance on the challenge website. Finally, we especially thank the NAACL Workshop chairs Prof. Bhavana Dalvi, Prof. Mamoru Komachi, and Prof. Michel Galley for helping us on all the complicated logistics for this year's online version.

We hope you enjoy the workshop!

### **Organizers**

Lun-Wei Ku, Academia Sincia, Taiwan

Cheng-Te Li, National Cheng Kung University, Taiwan

### **Challenge Organizers**

Lun-Wei Ku, Academia Sincia, Taiwan

Hong-Han Shuai, National Yang Ming Chiao Tung University, Taiwan

Ming Sun, Facebook



### **Organizers:**

Lun-Wei Ku, National Yang Ming Chiao Tung University  
Chenge-Te Li, National Cheng Kung University

### **Organizers of Fake-EmoReact Challenge:**

Lun-Wei Ku, Academia Sinica  
Hong-Han Shuai, Academia Sinica  
Ming Sun, Facebook

### **Program Committee:**

Khalid Al Khatib, Leipzig University  
Milad Alshomary, Paderborn University  
Silvio Amir, Northeastern University  
Guozhen An, City University of New York  
Sabine Bergler, Concordia University  
Laura Biester, University of Michigan  
Victoria Bobicev, Technical University of Moldova  
Caroline Brun, Institut NeuroMyoGène, Lyon  
Erik Cambria, Nanyang Technological University  
Paula Carvalho, UT Austin, Portugal CoLab  
Yung-Chun Chang, Taipei Medical University  
Yue Chen, Indiana University  
Zhuang Chen, WuHan University  
Hai Leong Chieu, DSO National Laboratories  
Patricia Chiril, Paul Sabatier University  
Oana Cocarascu, King's College London  
Danilo Croce, University of Roma, Tor Vergata  
Lei Cui, Microsoft Research  
Daniel Dakota, Uppsala University  
Pradipto Das, Rakuten USA  
MeiXing Dong, University of Michigan  
Rory Duthie, University of Dundee  
Wassim El-Hajj, American University of Beirut  
Elisabetta Fersini, University of Milano-Bicocca  
Lucie Flek, Philipps University of Marburg, TU Darmstadt  
Andrea Galassi, Università di Bologna  
Aparna Garimella, Adobe Inc  
Alexander Gelbukh, Instituto Politécnico Nacional  
Marco Guerini, Fondazione Bruno Kessler  
Tunga Güngör, Bogazici University  
Loitongbam Gyanendro Singh, IIT Guwahati  
Hatem Haddad, iCompass  
Chenyang Huang, University of Alberta  
Hen-Hsen Huang, National Chengchi University  
Kokil Jaidka, National University of Singapore  
Charles Jochim, IBM Research - Europe  
Xincheng Ju, Soochow University  
Pallika Kanani, Oracle Labs

Tsung-Ting Kuo, University of California San Diego  
 Yang Li, Northwestern Polytechnical University  
 Yingjie Li, University of Illinois at Chicago  
 Yang Li, Northeast Electric Power University  
 Peiqin Lin, Sun Yat-sen University  
 Chuan-Jie Lin, National Taiwan Ocean University  
 Marco Lippi, Università di Modena e Reggio Emilia  
 Pengfei Liu, SpeechX Limited  
 Avinash Madasu, Samsung R & D Institute Bangalore  
 Eugenio Martínez-Cámara, Universidad de Granada  
 Bruno Martins, University of Lisbon  
 Sahisnu Mazumder, University of Illinois at Chicago  
 Manuel Montes, INAOE, Mexico  
 Véronique MORICEAU, IRIT, Université de Toulouse  
 Hamdy Mubarak, Qatar Computing Research Institute QCRI  
 Nona Naderi, Bibliomics and Text Mining Group  
 Jose Ochoa-Luna, Universidad Católica San Pablo  
 Endang Wahyu Pamungkas, Università Degli Studi di Torino  
 Haris Papageorgiou, ATHENA Research and Innovation Center  
 Georgios Petasis, National Center for Scientific Research Demokritos  
 Omid Rohanian, University of Oxford  
 Mohammad Salameh, University of Alberta  
 Annika Marie Schoene, University of Manchester  
 Boaz Shmueli, National Tsing Hua University and Institute of Information Science, Academia Sinica  
 Abu Awal Md Shoeb, Rutgers University  
 Priyanka Sinha, Tata Consultancy Services  
 Irena Spasic, Cardiff University  
 Xavier Tannier, Sorbonne Université, INSERM, LIMICS  
 Paolo Torrioni, University of Bologna  
 Amine Trabelsi, University of Alberta  
 Enrica Troiano, University of Stuttgart  
 Paola Velardi, Sapienza University of Roma  
 Jingjing Wang, Tsinghua University  
 Hao Wang, Southwest Jiaotong University  
 Ingmar Weber, Qatar Computing Research Institute QCRI  
 Steven Wilson, The University of Edinburgh  
 Zhen Wu, Nanjing University  
 Shih-Hung Wu, Chaoyang University of Technology  
 Frank Xing, Nanyang Technological University  
 Lu Xu, Singapore University of Technology and Design  
 Jun Yang, Nanjing University  
 Zixiaofan Yang, Columbia University  
 Diyi Yang, Georgia Institute of Technology  
 Liang-Chih Yu, Yuan Ze University  
 Zhe Zhang, IBM  
 Lei Zhang, University of Illinois at Chicago  
 Bowen Zhang, Harbin Institute of Technology  
 Dongyu Zhang, Dalian University of Technology  
 Qi Zhang, Fudan University  
 Fei Zhao, Nanjing University

## Table of Contents

<i>Analysis of Nuanced Stances and Sentiment Towards Entities of US Politicians through the Lens of Moral Foundation Theory</i>	
Shamik Roy and Dan Goldwasser .....	1
<i>Content-based Stance Classification of Tweets about the 2020 Italian Constitutional Referendum</i>	
Marco Di Giovanni and Marco Brambilla .....	14
<i>A Case Study of In-House Competition for Ranking Constructive Comments in a News Service</i>	
Hayato Kobayashi, Hiroaki Taguchi, Yoshimune Tabuchi, Chahine Koleejan, Ken Kobayashi, Soichiro Fujita, Kazuma Murao, Takeshi Masuyama, Taichi Yatsuka, Manabu Okumura and Satoshi Sekine ..	24
<i>Quantifying the Effects of COVID-19 on Restaurant Reviews</i>	
Ivy Cao, Zizhou Liu, Giannis Karamanolakis, Daniel Hsu and Luis Gravano .....	36
<i>Assessing Cognitive Linguistic Influences in the Assignment of Blame</i>	
Karen Zhou, Ana Smith and Lillian Lee .....	61
<i>Evaluating Deception Detection Model Robustness To Linguistic Variation</i>	
Maria Glenski, Ellyn Ayton, Robin Cosbey, Dustin Arendt and Svitlana Volkova .....	70
<i>Reconsidering Annotator Disagreement about Racist Language: Noise or Signal?</i>	
Savannah Larimore, Ian Kennedy, Breon Haskett and Alina Arseniev-Koehler .....	81
<i>Understanding and Interpreting the Impact of User Context in Hate Speech Detection</i>	
Edoardo Mosca, Maximilian Wich and Georg Groh .....	91
<i>Self-Contextualized Attention for Abusive Language Identification</i>	
Horacio Jarquín-Vásquez, Hugo Jair Escalante and Manuel Montes .....	103
<i>Unsupervised Domain Adaptation in Cross-corpora Abusive Language Detection</i>	
Tulika Bose, Irina Illina and Dominique Fohr .....	113
<i>Using Noisy Self-Reports to Predict Twitter User Demographics</i>	
Zach Wood-Doughty, Paiheng Xu, Xiao Liu and Mark Dredze .....	123
<i>PANDORA Talks: Personality and Demographics on Reddit</i>	
Matej Gjurković, Mladen Karan, Iva Vukojević, Mihaela Bošnjak and Jan Snajder .....	138
<i>Room to Grow: Understanding Personal Characteristics Behind Self Improvement Using Social Media</i>	
MeiXing Dong, Xueming Xu, Yiwei Zhang, Ian Stewart and Rada Mihalcea .....	153
<i>Mitigating Temporal-Drift: A Simple Approach to Keep NER Models Crisp</i>	
Shuguang Chen, Leonardo Neves and Tamar Solorio .....	163
<i>Jujeop: Korean Puns for K-pop Stars on Social Media</i>	
Soyoung Oh, JISU KIM, Seungpeel Lee and Eunil Park .....	170
<i>Identifying Distributional Perspectives from Colingual Groups</i>	
Yufei Tian, Tuhin Chakrabarty, Fred Morstatter and Nanyun Peng .....	178





# Workshop Program

**June 10, 2021**

**9:00–9:10**     *Opening*

9:10–10:00     *Keynote Speech 1*  
Dan Goldwasser (Purdue University)

**10:00–10:10**     **Coffee Break**

**10:10–11:10**     **Technical Session 1: Sentiment, Reviews and Comments**

*Analysis of Nuanced Stances and Sentiment Towards Entities of US Politicians through the Lens of Moral Foundation Theory*

Shamik Roy and Dan Goldwasser

*Content-based Stance Classification of Tweets about the 2020 Italian Constitutional Referendum*

Marco Di Giovanni and Marco Brambilla

*A Case Study of In-House Competition for Ranking Constructive Comments in a News Service*

Hayato Kobayashi, Hiroaki Taguchi, Yoshimune Tabuchi, Chahine Koleejan, Ken Kobayashi, Soichiro Fujita, Kazuma Murao, Takeshi Masuyama, Taichi Yatsuka, Manabu Okumura and Satoshi Sekine

*Quantifying the Effects of COVID-19 on Restaurant Reviews*

Ivy Cao, Zizhou Liu, Giannis Karamanolakis, Daniel Hsu and Luis Gravano

**June 10, 2021 (continued)**

**11:10–11:40 Technical Session 2: Inappropriate Language (1)**

*Assessing Cognitive Linguistic Influences in the Assignment of Blame*

Karen Zhou, Ana Smith and Lillian Lee

*Evaluating Deception Detection Model Robustness To Linguistic Variation*

Maria Glenski, Ellyn Ayton, Robin Cosbey, Dustin Arendt and Svitlana Volkova

**11:40–12:40 Lunch Break, Poster and Fake-EmoReact Challenge**

12:40–13:40 *Keynote Speech 2*

Tim Weninger (University of Notre Dame)

**13:40–14:40 Technical Session 3: Inappropriate Language (2)**

*Reconsidering Annotator Disagreement about Racist Language: Noise or Signal?*

Savannah Larimore, Ian Kennedy, Breon Haskett and Alina Arseniev-Koehler

*Understanding and Interpreting the Impact of User Context in Hate Speech Detection*

Edoardo Mosca, Maximilian Wich and Georg Groh

*Self-Contextualized Attention for Abusive Language Identification*

Horacio Jarquín-Vásquez, Hugo Jair Escalante and Manuel Montes

*Unsupervised Domain Adaptation in Cross-corpora Abusive Language Detection*

Tulika Bose, Irina Illina and Dominique Fohr

**June 10, 2021 (continued)**

**14:40–14:50 Coffee Break**

**14:50–15:35 Technical Session 4: Personality and Demographics**

*Using Noisy Self-Reports to Predict Twitter User Demographics*

Zach Wood-Doughty, Paiheng Xu, Xiao Liu and Mark Dredze

*PANDORA Talks: Personality and Demographics on Reddit*

Matej Gjurković, Mladen Karan, Iva Vukojević, Mihaela Bošnjak and Jan Snajder

*Room to Grow: Understanding Personal Characteristics Behind Self Improvement Using Social Media*

MeiXing Dong, Xueming Xu, Yiwei Zhang, Ian Stewart and Rada Mihalcea

**15:35–16:20 Technical Session 5: Contexts and Perspectives**

*Mitigating Temporal-Drift: A Simple Approach to Keep NER Models Crisp*

Shuguang Chen, Leonardo Neves and Thamar Solorio

*Jujeop: Korean Puns for K-pop Stars on Social Media*

Soyoung Oh, JISU KIM, Seungpeel Lee and Eunil Park

*Identifying Distributional Perspectives from Colingual Groups*

Yufei Tian, Tuhin Chakrabarty, Fred Morstatter and Nanyun Peng

**16:20–16:30 Closing**

