

COLING

**International Conference on  
Computational Linguistics**

**Proceedings of the Conference and Workshops**

COLING

Volume 29 (2022), No. 4

**Proceedings of 8th Workshop on Noisy User-generated Text  
(W-NUT 2022)**

**The 29th International Conference on  
Computational Linguistics**

October 12 - 17, 2022  
Gyeongju, Republic of Korea

Copyright of each paper stays with the respective authors (or their employers).

ISSN 2951-2093

## Introduction

The W-NUT 2022 workshop focuses on a core set of natural language processing tasks on top of noisy user-generated text, such as that found on social media, web forums and online reviews. Recent years have seen a significant increase of interest in these areas. The internet has democratized content creation leading to an explosion of informal user-generated text, publicly available in electronic format, motivating the need for NLP on noisy text to enable new data analytics applications.

We have received 39 main workshop submissions (22 long and 17 short papers). The workshop will be held in hybrid in-person and virtual modes. We have two invited speakers Yulia Tsvetkov (University of Washington) and David Jurgens (University of Michigan) who will talk about their work. We're very thankful to have them in our workshop.

We have the best paper award(s) sponsored by Megagon Labs this year, for which we are thankful. We would like to thank the Program Committee members who reviewed the papers. We would also like to thank the workshop participants.

Wei Xu, Alan Ritter, Tim Baldwin and Afshin Rahimi  
Co-Organizers

**Organizers:**

Wei Xu, Ohio State University  
Alan Ritter, Ohio State University  
Tim Baldwin, University of Melbourne  
Afshin Rahimi, University of Queensland

**Program Committee:**

Muhammad Abdul-Mageed (University of British Columbia)  
Željko Agić (Corti)  
Sweta Agrawal (University of Maryland)  
Gustavo Aguilar (Amazon)  
Hamed Alhoori (Northern Illinois University)  
Emily Allaway (Columbia University)  
Hadi Amiri (University of Massachusetts, Lowell)  
Antonios Anastasopoulos (George Mason University)  
Maria Antoniak (Cornell University)  
Rahul Aralikkatte (University of Copenhagen)  
Eiji Aramaki (NAIST)  
Pepa Atanasova (University of Copenhagen)  
Ashutosh Baheti (Georgia Institute of Technology)  
JinYeong Bak (SungKyunKwan University)  
Kalika Bali (Microsoft Research)  
Francesco Barbieri (Snap)  
Elisa Bassignana (IT University of Copenhagen)  
Adrian Benton (JHU)  
Eduardo Blanco (University of North Texas)  
Marcel Bollmann (Jönköping University)  
Marco Brambilla (Politecnico di Milano)  
Julian Brooke (University of British Columbia)  
Cornelia Caragea (University of Illinois at Chicago)  
Tuhin Chakrabarty (Columbia University)  
Tanmoy Chakraborty (Indraprastha Institute of Information Technology, Delhi)  
Ilias Chalkidis (University of Copenhagen)  
Sihao Chen (University of Pennsylvania)  
Colin Cherry (Google)  
Dhivya Chinnappa (Thomson Reuters)  
Monojit Choudhury (Microsoft Research)  
Zewei Chu (University of Chicago)  
Manuel R. Ciosici (IT University of Copenhagen)  
Çağrı Çöltekin (University of Tübingen)  
Danilo Croce (University of Rome)  
Marina Danilevsky (IBM Research)  
Pradipto Das (Rakuten Institute of Technology)  
A. Seza Doğruöz (Universiteit Gent)  
Xinya Du (Cornell University)  
Heba Elfardy (Amazon)  
Mai ElSherief (University of California, San Diego)

Micha Elsner (Ohio State University)  
Alexander Fabbri (Yale University)  
Manaal Faruqui (Google)  
Song Feng (IBM Research)  
Yansong Feng (Peking University)  
Francis Ferraro (University of Maryland, Baltimore Countys)  
Catherine Finegan-Dollak (IBM Research)  
Lucie Flek (University of Marburg)  
Lisheng Fu (Amazon)  
Yoshinari Fujinuma (University of Colorado, Boulder)  
Wei Gao (Singapore Management University)  
Sahil Garg (University of Southern California)  
Dan Garrette (Google)  
Spandana Gella (Amazon)  
Tirthankar Ghosal (Charles University)  
Dan Goldwasser (Purdue University)  
Amit Goyal (Amazon)  
Yvette Graham (Dublin City University)  
Chulaka Gunasekara (IBM Research)  
Cathal Gurrin (Dublin City University)  
Xiaochuang Han (Carnegie Mellon University)  
Jonathan Herzig (Tel-Aviv University)  
Jack Hessel (AI2)  
Md Mosharaf Hossain (University of North Texas)  
Diana Inkpen (University of Ottawa)  
Kokil Jaidka (National University of Singapore)  
Yangfeng Ji (University of Virginia)  
Jing Jiang (Singapore Management University)  
Nanjiang Jiang (Ohio State University)  
Chao Jiang (Georgia Institute of Technology)  
Lifeng Jin (Ohio State University)  
Ishan Jindal (IBM Research)  
Kristen Johnson (Michigan State University)  
Gareth Jones (Dublin City University)  
David Jurgens (University of Michigan)  
Preethi Jyothi (IIT Bombay)  
Katharina Kann (University of Colorado, Boulder)  
Ashkan Kazemi (University of Michigan)  
Ashique KhudaBukhsh (Carnegie Mellon University)  
Gunhee Kim (Seoul National University)  
Bennett Kleinberg (Tilburg University)  
Roman Klinger (University of Stuttgart)  
Zornitsa Kozareva (Facebook)  
Reno Kriz (University of Pennsylvania)  
Vivek Kulkarni (Stanford University)  
Jonathan Kummerfeld (University of Michigan)  
Tsung-Ting Kuo (University of California, San Diego)  
Vasileios Lamos (University College London)  
Wuwei Lan (Ohio State University)  
Jiwei Li (ShannonAI)  
Jessy Junyi Li (University of Texas Austin)

Jing Li (Hong Kong Polytechnic University)  
Yitong Li (University of Melbourne)  
Kwan Hui Lim (Singapore University of Technology and Design)  
Nut Limsopatham (Microsoft Research)  
Lucy Lin (University of Washington)  
Fei Liu (University of Melbourne)  
Yiqun Liu (Tsinghua University)  
Nikola Ljubešić (Jožef Stefan Institute)  
Mounica Maddela (Georgia Institute of Technology)  
Bodhisattwa Prasad Majumder (University of California, San Diego)  
David Mimno (Cornell University)  
Shachar Mirkin (Digimind)  
Yasuhide Miura (Fuji/Xerox)  
Manuel Montes (National Institute of Astrophysics, Mexico)  
Ahmed Mourad (RMIT University)  
Maximilian Mozes (University College London)  
Hamdy Mubarak (Qatar Computing Research Institute)  
Graham Mueller (Leidos)  
Animesh Mukherjee (IIT Kharagpur)  
Maria Nadejde (Grammarly)  
Preslav Nakov (Qatar Computing Research Institute)  
Guenter Neumann (German Research Center for Artificial Intelligence)  
Vincent Ng (University of Texas at Dallas)  
Thien Huu Nguyen (University of Oregon)  
Dat Quoc Nguyen (VinAI Research)  
Eric Nichols (Honda Research Institute)  
Brendan O'Connor (University of Massachusetts, Amherst)  
Alice Oh (KAIST)  
Naoaki Okazaki (Tohoku University)  
Naoki Otani (CMU)  
Symeon Papadopoulos (CERTH-ITI)  
Yuval Pinter (Georgia Tech)  
Barbara Plank (IT University of Copenhagen)  
Matt Post (Johns Hopkins University)  
Vinodkumar Prabhakaran (Stanford University)  
Daniel Preoțiuc-Pietro (Bloomberg)  
Dianna Radpour (University of Colorado Boulder)  
Preethi Raghavan (IBM Research)  
Afshin Rahimi (University of Queensland)  
Sudha Rao (Microsoft Research)  
Hannah Rashkin (Google)  
Sravana Reddy (ASAPP)  
Roi Reichart (Technion - Israel Institute of Technology.)  
Adithya Renduchintala (Facebook AI)  
Anthony Rios (The University of Texas at San Antonio)  
Paolo Rosso (Universitat Politècnica de València)  
Alla Rozovskaya (City University of New York)  
Mirco Schoenfeld (University of Bayreuth)  
Djamé Seddah (University Paris-Sorbonne)  
Ori Shapira (Bar-Ilan University)  
Ashish Sharma (University of Washington)

Dan Simonson (BlackBoiler)  
Kevin Small (Amazon)  
Xingyi Song (University of Sheffield)  
Andreas Spitz (University of Konstanz)  
Richard Sproat (Google)  
Gabriel Stanovsky (Allen Institute for Artificial Intelligence)  
Ian Stewart (University of Michigan)  
Sara Stymne (Uppsala University)  
Danae Sánchez Villegas (University of Sheffield)  
Zeeraq Talat (University of Sheffield)  
Zhiyang Teng (Westlake University)  
James Thorne (University of Cambridge)  
Marc Tomlinson (Language Computer Corporation)  
Sara Tonelli (FBK)  
Rob van der Goot (University of Groningen)  
Vasudeva Varma (IIIT Hyderabad)  
Daniel Varab (IT University of Copenhagen)  
Olga Vechtomova (University of Waterloo)  
Rob Voigt (Northwestern University)  
Soroush Vosoughi (Dartmouth University)  
Thanh Vu (Oracle)  
Xiaojun Wan (Peking University)  
Hong Wei (University of Maryland)  
Zhongyu Wei (Fudan University)  
Roman Yangarber (University of Helsinki)  
Ziyu Yao (George Mason University)  
Ning Yu (Leidos)  
Nasser Zalmout (Amazon)  
Marcos Zampieri (Rochester Institute of Technology)  
Vicky Zayats (University of Washington)  
Chiyu Zhang (University of British Columbia)  
Xiao Cosmo Zhang (Amazon)  
Mike Zhang (IT University of Copenhagen)  
Ayah Zirikly (Johns Hopkins University)  
Shi Zong (Nanjing University)  
Hamid Beigy (Sharif University of Technology)  
Soroush Vosoughi (Dartmouth College)

**Invited Speakers:**

Yulia Tsvetkov (University of Washington)  
David Jurgens (University of Michigan)



## Table of Contents

<i>Changes in Tweet Geolocation over Time: A Study with Carmen 2.0</i> Jingyu Zhang, Alexandra DeLucia and Mark Dredze .....	1
<i>Extracting Mathematical Concepts from Text</i> Jacob Collard, Valeria de Paiva, Brendan Fong and Eswaran Subrahmanian .....	15
<i>Data-driven Approach to Differentiating between Depression and Dementia from Noisy Speech and Language Data</i> Malikeh Ehghaghi, Frank Rudzicz and Jekaterina Novikova .....	24
<i>Cross-Dialect Social Media Dependency Parsing for Social Scientific Entity Attribute Analysis</i> Chloe Eggleston and Brendan O'Connor .....	38
<i>Impact of Environmental Noise on Alzheimer's Disease Detection from Speech: Should You Let a Baby Cry?</i> Jekaterina Novikova .....	51
<i>Exploring Multimodal Features and Fusion Strategies for Analyzing Disaster Tweets</i> Raj Pranesh .....	62
<i>NTULM: Enriching Social Media Text Representations with Non-Textual Units</i> Jinning Li, Shubhanshu Mishra, Ahmed El-Kishky, Sneha Mehta and Vivek Kulkarni .....	69
<i>Robust Candidate Generation for Entity Linking on Short Social Media Texts</i> Liam Hebert, Raheleh Makki, Shubhanshu Mishra, Hamidreza Saghir, Anusha Kamath and Yuval Merhav .....	83
<i>TransPOS: Transformers for Consolidating Different POS Tagset Datasets</i> Alex Li, Ilyas Bankole-Hameed, Ranadeep Singh, Gabriel Ng and Akshat Gupta .....	90
<i>An Effective, Performant Named Entity Recognition System for Noisy Business Telephone Conversation Transcripts</i> Xue-Yong Fu, Cheng Chen, Md Tahmid Rahman Laskar, Shashi Bhushan TN and Simon Corston-Oliver .....	96
<i>Leveraging Semantic and Sentiment Knowledge for User-Generated Text Sentiment Classification</i> Jawad Khan, Niaz Ahmad, Aftab Alam and Youngmoon Lee .....	101
<i>An Emotional Journey: Detecting Emotion Trajectories in Dutch Customer Service Dialogues</i> Sofie Labat, Amir Hadifar, Thomas Demeester and Veronique Hoste .....	106
<i>Supervised and Unsupervised Evaluation of Synthetic Code-Switching</i> Evgeny Orlov and Ekaterina Artemova .....	113
<i>ArabGend: Gender Analysis and Inference on Arabic Twitter</i> Hamdy Mubarak, Shammur Absar Chowdhury and Firoj Alam .....	124
<i>Automatic Identification of 5C Vaccine Behaviour on Social Media</i> Ajay Hemanth Sampath Kumar, Aminath Shausan, Gianluca Demartini and Afshin Rahimi ...	136

<i>Automatic Extraction of Structured Mineral Drillhole Results from Unstructured Mining Company Reports</i>	
Adam Dimeski and Afshin Rahimi .....	147
<i>"Kanglish alli names!" Named Entity Recognition for Kannada-English Code-Mixed Social Media Data</i>	
Sumukh S and Manish Shrivastava .....	154
<i>Span Extraction Aided Improved Code-mixed Sentiment Classification</i>	
Ramaneswaran S, Sean Benhur and Sreyan Ghosh .....	162
<i>AdBERT: An Effective Few Shot Learning Framework for Aligning Tweets to Superbowl Advertisements</i>	
Debarati Das, Roopana Chenchu, Maral Abdollahi, Jisu Huh and Jaideep Srivastava .....	171
<i>Increasing Robustness for Cross-domain Dialogue Act Classification on Social Media Data</i>	
Marcus Vielsted, Nikolaj Wallenius and Rob van der Goot .....	180
<i>Disfluency Detection for Vietnamese</i>	
Mai Dao, Think Hung Truong and Dat Quoc Nguyen .....	194
<i>A multi-level approach for hierarchical Ticket Classification</i>	
Matteo Marcuzzo, Alessandro Zangari, Michele Schiavinato, Lorenzo Giudice, Andrea Gasparetto and Andrea Albarelli .....	201
<i>Towards better structured and less noisy Web data: Oscar with Register annotations</i>	
Veronika Laippala, Anna Salmela, Samuel Rönqvist, Alham Fikri Aji, Li-Hsin Chang, Asma Dhifallah, Larissa Goulart, Henna Kortelainen, Marc Pàmies, Deise Prina Dutra, Valtteri Skantsi, Lintang Sutawika and Sampo Pyysalo .....	215
<i>True or False? Detecting False Information on Social Media Using Graph Neural Networks</i>	
Samyo Rode-Hasinger, Anna Kruspe and Xiao Xiang Zhu .....	222
<i>Analyzing the Real Vulnerability of Hate Speech Detection Systems against Targeted Intentional Noise</i>	
Piush Aggarwal and Torsten Zesch .....	230

# Conference Program

**Saturday 16 October 2022**

**9:00 AM–11:19 AM Morning Session**

Welcome

Invited Talk 1

*Changes in Tweet Geolocation over Time: A Study with Carmen 2.0*

Jingyu Zhang, Alexandra DeLucia and Mark Dredze

*Extracting Mathematical Concepts from Text*

Jacob Collard, Valeria de Paiva, Brendan Fong and Eswaran Subrahmanian

*Data-driven Approach to Differentiating between Depression and Dementia from Noisy Speech and Language Data*

Malikeh Ehghaghi, Frank Rudzicz and Jekaterina Novikova

*Cross-Dialect Social Media Dependency Parsing for Social Scientific Entity Attribute Analysis*

Chloe Eggleston and Brendan O'Connor

*Impact of Environmental Noise on Alzheimer's Disease Detection from Speech: Should You Let a Baby Cry?*

Jekaterina Novikova

*Exploring Multimodal Features and Fusion Strategies for Analyzing Disaster Tweets*

Raj Pranesh

*NTULM: Enriching Social Media Text Representations with Non-Textual Units*

Jinning Li, Shubhanshu Mishra, Ahmed El-Kishky, Sneha Mehta and Vivek Kulka-rni

*Robust Candidate Generation for Entity Linking on Short Social Media Texts*

Liam Hebert, Raheleh Makki, Shubhanshu Mishra, Hamidreza Saghir, Anusha Ka-math and Yuval Merhav

*TransPOS: Transformers for Consolidating Different POS Tagset Datasets*

Alex Li, Ilyas Bankole-Hameed, Ranadeep Singh, Gabriel Ng and Akshat Gupta

**Saturday 16 October 2022 (continued)**

*An Effective, Performant Named Entity Recognition System for Noisy Business Telephone Conversation Transcripts*

Xue-Yong Fu, Cheng Chen, Md Tahmid Rahman Laskar, Shashi Bhushan TN and Simon Corston-Oliver

Gather.Town social event

**12:20**

**PM–3:40 PM Afternoon Session**

Invited Talk 2

*Leveraging Semantic and Sentiment Knowledge for User-Generated Text Sentiment Classification*

Jawad Khan, Niaz Ahmad, Aftab Alam and Youngmoon Lee

*An Emotional Journey: Detecting Emotion Trajectories in Dutch Customer Service Dialogues*

Sofie Labat, Amir Hadifar, Thomas Demeester and Veronique Hoste

*Supervised and Unsupervised Evaluation of Synthetic Code-Switching*

Evgeny Orlov and Ekaterina Artemova

*ArabGend: Gender Analysis and Inference on Arabic Twitter*

Hamdy Mubarak, Shammur Absar Chowdhury and Firoj Alam

*Automatic Identification of 5C Vaccine Behaviour on Social Media*

Ajay Hemanth Sampath Kumar, Aminath Shausan, Gianluca Demartini and Afshin Rahimi

*Automatic Extraction of Structured Mineral Drillhole Results from Unstructured Mining Company Reports*

Adam Dimeski and Afshin Rahimi

*"Kanglish alli names!" Named Entity Recognition for Kannada-English Code-Mixed Social Media Data*

Sumukh S and Manish Shrivastava

Gather.Town social event

**Saturday 16 October 2022 (continued)**

*Span Extraction Aided Improved Code-mixed Sentiment Classification*

Ramaneswaran S, Sean Benhur and Sreyan Ghosh

*AdBERT: An Effective Few Shot Learning Framework for Aligning Tweets to Super-bowl Advertisements*

Debarati Das, Roopana Chenchu, Maral Abdollahi, Jisu Huh and Jaideep Srivastava

*Increasing Robustness for Cross-domain Dialogue Act Classification on Social Media Data*

Marcus Vielsted, Nikolaj Wallenius and Rob van der Goot

*Disfluency Detection for Vietnamese*

Mai Dao, Thinh Hung Truong and Dat Quoc Nguyen

*A multi-level approach for hierarchical Ticket Classification*

Matteo Marcuzzo, Alessandro Zangari, Michele Schiavinato, Lorenzo Giudice, Andrea Gasparetto and Andrea Albarelli

*Towards better structured and less noisy Web data: Oscar with Register annotations*

Veronika Laippala, Anna Salmela, Samuel Rönqvist, Alham Fikri Aji, Li-Hsin Chang, Asma Dhifallah, Larissa Goulart, Henna Kortelainen, Marc Pàmies, Deise Prina Dutra, Valteri Skantsi, Lintang Sutawika and Sampo Pyysalo

*True or False? Detecting False Information on Social Media Using Graph Neural Networks*

Samyo Rode-Hasinger, Anna Kruspe and Xiao Xiang Zhu

*Analyzing the Real Vulnerability of Hate Speech Detection Systems against Targeted Intentional Noise*

Piush Aggarwal and Torsten Zesch

Gather.Town social event

Best paper award and closing

