

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

Gyeongju, South Korea
12-17 October 2022

ISBN: 978-1-7138-6200-0

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571



Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2022) by the Association for Computational Linguistics
All rights reserved.

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

Printed with permission by Curran Associates, Inc. (2023)

For permission requests, please contact the Association for Computational Linguistics
at the address below.

Association for Computational Linguistics
209 N. Eighth Street
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006

Fax: 1-570-476-0860

acl@aclweb.org

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2633
Email: curran@proceedings.com
Web: www.proceedings.com

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, Thinh 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 Rönnqvist, 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