

Sixth Workshop on Noisy User-generated Text (W-NUT 2020)

Online
19 November 2020

ISBN: 978-1-7138-2006-2

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© (2020) by the Association for Computational Linguistics
All rights reserved.

Printed with permission by Curran Associates, Inc. (2021)

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>May I Ask Who's Calling? Named Entity Recognition on Call Center Transcripts for Privacy Law Compliance</i>	
Micaela Kaplan	1
<i>"Did you really mean what you said?" : Sarcasm Detection in Hindi-English Code-Mixed Data using Bilingual Word Embeddings</i>	
Akshita Aggarwal, Anshul Wadhawan, Anshima Chaudhary and Kavita Maurya	7
<i>Noisy Text Data: Achilles' Heel of BERT</i>	
Ankit Kumar, Piyush Makhija and Anuj Gupta	16
<i>Determining Question-Answer Plausibility in Crowdsourced Datasets Using Multi-Task Learning</i>	
Rachel Gardner, Maya Varma, Clare Zhu and Ranjay Krishna	22
<i>Combining BERT with Static Word Embeddings for Categorizing Social Media</i>	
Israa Alghanmi, Luis Espinosa Anke and Steven Schockaert	28
<i>Enhanced Sentence Alignment Network for Efficient Short Text Matching</i>	
Zhe Hu, Zuohui Fu, Cheng Peng and Weiwei Wang	34
<i>PHINC: A Parallel Hinglish Social Media Code-Mixed Corpus for Machine Translation</i>	
Vivek Srivastava and Mayank Singh	41
<i>Cross-lingual sentiment classification in low-resource Bengali language</i>	
Salim Sazed	50
<i>The Non-native Speaker Aspect: Indian English in Social Media</i>	
Rupak Sarkar, Sayantan Mahinder and Ashiqur KhudaBukhsh	61
<i>Sentence Boundary Detection on Line Breaks in Japanese</i>	
Yuta Hayashibe and Kensuke Mitsuzawa	71
<i>Non-ingredient Detection in User-generated Recipes using the Sequence Tagging Approach</i>	
Yasuhiro Yamaguchi, Shintaro Inuzuka, Makoto Hiramatsu and Jun Harashima	76
<i>Generating Fact Checking Summaries for Web Claims</i>	
Rahul Mishra, Dhruv Gupta and Markus Leippold	81
<i>Intelligent Analyses on Storytelling for Impact Measurement</i>	
Koen Kicken, Tessa De Maesschalck, Bart Vanrumste, Tom De Keyser and Hee Reen Shim	91
<i>An Empirical Analysis of Human-Bot Interaction on Reddit</i>	
Ming-Cheng Ma and John P. Lalor	101
<i>Detecting Trending Terms in Cybersecurity Forum Discussions</i>	
Jack Hughes, Seth Aycock, Andrew Caines, Paula BATTERY and Alice Hutchings	107
<i>Service registration chatbot: collecting and comparing dialogues from AMT workers and service's users</i>	
Luca Molteni, Mittul Singh, Juho Leinonen, Katri Leino, Mikko Kurimo and Emanuele Della Valle	116
<i>Automated Assessment of Noisy Crowdsourced Free-text Answers for Hindi in Low Resource Setting</i>	
Dolly Agarwal, Somya Gupta and Nishant Baghel	122

<i>Punctuation Restoration using Transformer Models for Resource-Rich and -Poor Languages</i>	
Tanvirul Alam, Akib Khan and Firoj Alam	132
<i>Truecasing German user-generated conversational text</i>	
Yulia Grishina, Thomas Gueudre and Ralf Winkler	143
<i>Fine-Tuning MT systems for Robustness to Second-Language Speaker Variations</i>	
Md Mahfuz Ibn Alam and Antonios Anastasopoulos	149
<i>Impact of ASR on Alzheimer’s Disease Detection: All Errors are Equal, but Deletions are More Equal than Others</i>	
Aparna Balagopalan, Ksenia Shkaruta and Jekaterina Novikova	159
<i>Detecting Entailment in Code-Mixed Hindi-English Conversations</i>	
Sharanya Chakravarthy, Anjana Umapathy and Alan W Black	165
<i>Detecting Objectifying Language in Online Professor Reviews</i>	
Angie Waller and Kyle Gorman	171
<i>Annotation Efficient Language Identification from Weak Labels</i>	
Shriphani Palakodety and Ashiqur KhudaBukhsh	181
<i>Fantastic Features and Where to Find Them: Detecting Cognitive Impairment with a Subsequence Classification Guided Approach</i>	
Ben Eyre, Aparna Balagopalan and Jekaterina Novikova	193
<i>Quantifying the Evaluation of Heuristic Methods for Textual Data Augmentation</i>	
Omid Kashefi and Rebecca Hwa	200
<i>An Empirical Survey of Unsupervised Text Representation Methods on Twitter Data</i>	
Lili Wang, Chongyang Gao, Jason Wei, Weicheng Ma, Ruibo Liu and Soroush Vosoughi	209
<i>Civil Unrest on Twitter (CUT): A Dataset of Tweets to Support Research on Civil Unrest</i>	
Justin Sech, Alexandra DeLucia, Anna L Buczak and Mark Dredze	215
<i>Tweeki: Linking Named Entities on Twitter to a Knowledge Graph</i>	
Bahareh Harandizadeh and Sameer Singh	222
<i>Representation learning of writing style</i>	
Julien Hay, Bich-Lien Doan, Fabrice Popineau and Ouassim Ait Elhara	232
<i>"A Little Birdie Told Me ... " - Social Media Rumor Detection</i>	
Karthik Radhakrishnan, Tushar Kanakagiri, Sharanya Chakravarthy and Vidhisha Balachandran	244
<i>Paraphrase Generation via Adversarial Penalizations</i>	
Gerson Vizcarra and Jose Ochoa-Luna	249
<i>WNUT-2020 Task 1 Overview: Extracting Entities and Relations from Wet Lab Protocols</i>	
Jeniya Tabassum, Wei Xu and Alan Ritter	260
<i>IITKGP at W-NUT 2020 Shared Task-1: Domain specific BERT representation for Named Entity Recognition of lab protocol</i>	
Tejas Vaidhya and Ayush Kaushal	268

<i>PublishInCovid19 at WNUT 2020 Shared Task-1: Entity Recognition in Wet Lab Protocols using Structured Learning Ensemble and Contextualised Embeddings</i>	
Janvijay Singh and Anshul Wadhawan	273
<i>Big Green at WNUT 2020 Shared Task-1: Relation Extraction as Contextualized Sequence Classification</i>	
Chris Miller and Soroush Vosoughi	281
<i>WNUT 2020 Shared Task-1: Conditional Random Field(CRF) based Named Entity Recognition(NER) for Wet Lab Protocols</i>	
Kaushik Acharya	286
<i>mgsohrab at WNUT 2020 Shared Task-1: Neural Exhaustive Approach for Entity and Relation Recognition Over Wet Lab Protocols</i>	
Mohammad Golam Sohrab, Anh-Khoa Duong Nguyen, Makoto Miwa and Hiroya Takamura ..	290
<i>Fancy Man Launches Zippo at WNUT 2020 Shared Task-1: A Bert Case Model for Wet Lab Entity Extraction</i>	
Qingcheng Zeng, Xiaoyang Fang, Zhixin Liang and Haoding Meng	299
<i>BiTeM at WNUT 2020 Shared Task-1: Named Entity Recognition over Wet Lab Protocols using an Ensemble of Contextual Language Models</i>	
Julien Knafou, Nona Naderi, Jenny Copara, Douglas Teodoro and Patrick Ruch	305
<i>WNUT-2020 Task 2: Identification of Informative COVID-19 English Tweets</i>	
Dat Quoc Nguyen, Thanh Vu, Afshin Rahimi, Mai Hoang Dao, Linh The Nguyen and Long Doan	314
<i>TATL at WNUT-2020 Task 2: A Transformer-based Baseline System for Identification of Informative COVID-19 English Tweets</i>	
Anh Tuan Nguyen	319
<i>NHK_STRL at WNUT-2020 Task 2: GATs with Syntactic Dependencies as Edges and CTC-based Loss for Text Classification</i>	
Yuki Yasuda, Taichi Ishiwatari, Taro Miyazaki and Jun Goto	324
<i>NLP North at WNUT-2020 Task 2: Pre-training versus Ensembling for Detection of Informative COVID-19 English Tweets</i>	
Anders Giovanni Møller, Rob van der Goot and Barbara Plank	331
<i>Siva at WNUT-2020 Task 2: Fine-tuning Transformer Neural Networks for Identification of Informative Covid-19 Tweets</i>	
Siva Sai	337
<i>IITBH at WNUT-2020 Task 2: Exploiting the best of both worlds</i>	
Saichethan Reddy and Pradeep Biswal	342
<i>Phonemer at WNUT-2020 Task 2: Sequence Classification Using COVID Twitter BERT and Bagging Ensemble Technique based on Plurality Voting</i>	
Anshul Wadhawan	347
<i>CXP949 at WNUT-2020 Task 2: Extracting Informative COVID-19 Tweets - RoBERTa Ensembles and The Continued Relevance of Handcrafted Features</i>	
Calum Perrio and Harish Tayyar Madabushi	352

<i>InfoMiner at WNUT-2020 Task 2: Transformer-based Covid-19 Informative Tweet Extraction</i>	
Hansi Hettiarachchi and Tharindu Ranasinghe	359
<i>BANANA at WNUT-2020 Task 2: Identifying COVID-19 Information on Twitter by Combining Deep Learning and Transfer Learning Models</i>	
Tin Huynh, Luan Thanh Luan and Son T. Luu	366
<i>DATAMAFIA at WNUT-2020 Task 2: A Study of Pre-trained Language Models along with Regularization Techniques for Downstream Tasks</i>	
Ayan Sengupta	371
<i>UPennHLP at WNUT-2020 Task 2 : Transformer models for classification of COVID19 posts on Twitter</i>	
Arjun Magge, Varad Pimpalkhute, Divya Rallapalli, David Siguenza and Graciela Gonzalez-Hernandez	
378	
<i>UIT-HSE at WNUT-2020 Task 2: Exploiting CT-BERT for Identifying COVID-19 Information on the Twitter Social Network</i>	
Khiem Tran, Hao Phan, Kiet Nguyen and Ngan Luu Thuy Nguyen	383
<i>Emory at WNUT-2020 Task 2: Combining Pretrained Deep Learning Models and Feature Enrichment for Informative Tweet Identification</i>	
Yuting Guo, Mohammed Ali Al-Garadi and Abeed Sarker	388
<i>CSECU-DSG at WNUT-2020 Task 2: Exploiting Ensemble of Transfer Learning and Hand-crafted Features for Identification of Informative COVID-19 English Tweets</i>	
Fareen Tasneem, Jannatun Naim, Radiathun Tasnia, Tashin Hossain and Abu Nowshed Chy ...	394
<i>IRLab@IITBHU at WNUT-2020 Task 2: Identification of informative COVID-19 English Tweets using BERT</i>	
Supriya Chanda, Eshita Nandy and Sukomal Pal	399
<i>NutCracker at WNUT-2020 Task 2: Robustly Identifying Informative COVID-19 Tweets using Ensembling and Adversarial Training</i>	
Priyanshu Kumar and Aadarsh Singh	404
<i>DSC-IIT ISM at WNUT-2020 Task 2: Detection of COVID-19 informative tweets using RoBERTa</i>	
Sirigireddy Dhana Laxmi, Rohit Agarwal and Aman Sinha	409
<i>Linguist Geeks on WNUT-2020 Task 2: COVID-19 Informative Tweet Identification using Progressive Trained Language Models and Data Augmentation</i>	
Vasudev Awatramani and Anupam Kumar	414
<i>NLPRL at WNUT-2020 Task 2: ELMo-based System for Identification of COVID-19 Tweets</i>	
Rajesh Kumar Mundotiya, Rupjyoti Baruah, Bhavana Srivastava and Anil Kumar Singh	419
<i>SU-NLP at WNUT-2020 Task 2: The Ensemble Models</i>	
Kenan Fayoumi and Reyyan Yeniterzi	423
<i>IDSOU at WNUT-2020 Task 2: Identification of Informative COVID-19 English Tweets</i>	
Sora Ohashi, Tomoyuki Kajiwara, Chenhui Chu, Noriko Takemura, Yuta Nakashima and Hajime Nagahara	428
<i>ComplexDataLab at W-NUT 2020 Task 2: Detecting Informative COVID-19 Tweets by Attending over Linked Documents</i>	
Kellin Pelrine, Jacob Danovitch, Albert Orozco Camacho and Reihaneh Rabbany	434

<i>NEU at WNUT-2020 Task 2: Data Augmentation To Tell BERT That Death Is Not Necessarily Informative</i> Kumud Chauhan	440
<i>LynnyrdSkynnyrd at WNUT-2020 Task 2: Semi-Supervised Learning for Identification of Informative COVID-19 English Tweets</i> Abhilasha Sancheti, Kushal Chawla and Gaurav Verma.....	444
<i>NIT_COVID-19 at WNUT-2020 Task 2: Deep Learning Model RoBERTa for Identify Informative COVID-19 English Tweets</i> Jagadeesh M S and Alphonse P J A.....	450
<i>EdinburghNLP at WNUT-2020 Task 2: Leveraging Transformers with Generalized Augmentation for Identifying Informativeness in COVID-19 Tweets</i> Nickil Maveli	455
<i>#GCDH at WNUT-2020 Task 2: BERT-Based Models for the Detection of Informativeness in English COVID-19 Related Tweets</i> Hanna Varachkina, Stefan Ziehe, Tillmann Dönicke and Franziska Pannach	462
<i>Not-NUTs at WNUT-2020 Task 2: A BERT-based System in Identifying Informative COVID-19 English Tweets</i> Thai Hoang and Phuong Vu.....	466
<i>CIA_NITT at WNUT-2020 Task 2: Classification of COVID-19 Tweets Using Pre-trained Language Models</i> Yandrapati Prakash Babu and Rajagopal Eswari	471
<i>UET at WNUT-2020 Task 2: A Study of Combining Transfer Learning Methods for Text Classification with RoBERTa</i> Huy Dao Quang and Tam Nguyen Minh	475
<i>Dartmouth CS at WNUT-2020 Task 2: Fine tuning BERT for Tweet classification</i> Dylan Whang and Soroush Vosoughi	480
<i>SunBear at WNUT-2020 Task 2: Improving BERT-Based Noisy Text Classification with Knowledge of the Data domain</i> Linh Doan Bao, Viet Anh Nguyen and Quang Pham Huu	485
<i>ISWARA at WNUT-2020 Task 2: Identification of Informative COVID-19 English Tweets using BERT and FastText Embeddings</i> Wava Carissa Putri, Rani Aulia Hidayat, Isnaini Nurul Khasanah and Rahmad Mahendra	491
<i>COVCOR20 at WNUT-2020 Task 2: An Attempt to Combine Deep Learning and Expert rules</i> Ali Hürriyetoğlu, Ali Safaya, Osman Mutlu, Nelleke Oostdijk and Erdem Yörük	495
<i>TEST_POSITIVE at W-NUT 2020 Shared Task-3: Cross-task modeling</i> Chacha Chen, Chieh-Yang Huang, Yaqi Hou, Yang Shi, Enyan Dai and Jiaqi Wang.....	499
<i>imec-ETRO-VUB at W-NUT 2020 Shared Task-3: A multilabel BERT-based system for predicting COVID-19 events</i> Xiangyu Yang, Giannis Bekoulis and Nikos Deligiannis	505
<i>UCD-CS at W-NUT 2020 Shared Task-3: A Text to Text Approach for COVID-19 Event Extraction on Social Media</i> Congcong Wang and David Lillis	514

<i>Winners at W-NUT 2020 Shared Task-3: Leveraging Event Specific and Chunk Span information for Extracting COVID Entities from Tweets</i>	
Ayush Kaushal and Tejas Vaidhya	522
<i>HLTRI at W-NUT 2020 Shared Task-3: COVID-19 Event Extraction from Twitter Using Multi-Task Hop-field Pooling</i>	
Maxwell Weinzierl and Sanda Harabagiu	530