

W-NUT 2021

**The Seventh Workshop on
Noisy User-generated Text
(W-NUT 2021)**

Proceedings of the Conference

Nov 11, 2021
Online

©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

ISBN 978-1-954085-90-9

Introduction

The W-NUT 2021 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.

This year, in addition to the main workshop track, we have one shared task, MultiLexNorm: Multilingual Lexical Normalization. We have received 69 main workshop submissions (38 long and 31 short papers). The workshop will be held online and live in two different time zones (GMT- and GMT++). We have two invited speakers in GMT-, Zornitsa Kozareva (Google) Thamar Solorio (University of Houston), and two invited speakers in GMT++, Vivian Chen (National Taiwan University) and Zhiyuan Liu (Tsinghua University) 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 Bloomberg this year, for which we are thankful. We would like to thank the Program Committee members who reviewed the papers and the shared task organizers who enriched our workshop this year. 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 Faruqi (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)

Invited Speakers:

Zornitsa Kozareva (Google)
Thamar Solorio (University of Houston)
Vivian Chen (National Taiwan University)
Zhiyuan Liu (Tsinghua University)

Table of Contents

<i>Text Simplification for Comprehension-based Question-Answering</i> Tanvi Dadu, Kartikey Pant, Seema Nagar, Ferdous Barbhuiya and Kuntal Dey	1
<i>Finding the needle in a haystack: Extraction of Informative COVID-19 Danish Tweets</i> Benjamin Olsen and Barbara Plank	11
<i>Detecting Depression in Thai Blog Posts: a Dataset and a Baseline</i> Mika Hämäläinen, Pattama Patpong, Khalid Alnajjar, Niko Partanen and Jack Rueter	20
<i>Keyphrase Extraction with Incomplete Annotated Training Data</i> Yanfei Lei, Chunming Hu, Guanghui Ma and Richong Zhang	26
<i>Fine-grained Temporal Relation Extraction with Ordered-Neuron LSTM and Graph Convolutional Networks</i> Minh Tran Phu, Minh Van Nguyen and Thien Huu Nguyen	35
<i>Does It Happen? Multi-hop Path Structures for Event Factuality Prediction with Graph Transformer Networks</i> Duong Le and Thien Huu Nguyen	46
<i>Google-trickers, Yaminjeongeum, and Leetspeak: An Empirical Taxonomy for Intentionally Noisy User-Generated Text</i> Won Ik Cho and Soomin Kim	56
<i>Description-based Label Attention Classifier for Explainable ICD-9 Classification</i> Malte Feucht, Zhiliang Wu, Sophia Althammer and Volker Tresp	62
<i>A Text Editing Approach to Joint Japanese Word Segmentation, POS Tagging, and Lexical Normalization</i> Shohei Higashiyama, Masao Utiyama, Taro Watanabe and Eiichiro Sumita	67
<i>Intrinsic evaluation of language models for code-switching</i> Sik Feng Cheong, Hai Leong Chieu and Jing Lim	81
<i>Can images help recognize entities? A study of the role of images for Multimodal NER</i> Shuguang Chen, Gustavo Aguilar, Leonardo Neves and Thamar Solorio	87
<i>Perceived and Intended Sarcasm Detection with Graph Attention Networks</i> Joan Plepi and Lucie Flek	97
<i>Hierarchical Character Tagger for Short Text Spelling Error Correction</i> Mengyi Gao, Canran Xu and Peng Shi	106
<i>Common Sense Bias in Semantic Role Labeling</i> Heather Lent and Anders Søgaard	114
<i>PoliWAM: An Exploration of a Large Scale Corpus of Political Discussions on WhatsApp Messenger</i> Vivek Srivastava and Mayank Singh	120
<i>ParsTwiNER: A Corpus for Named Entity Recognition at Informal Persian</i> MohammadMahdi Aghajani, AliAkbar Badri and Hamid Beigy	131
<i>DreamDrug - A crowdsourced NER dataset for detecting drugs in darknet markets</i> Johannes Bogenasperger, Sven Schlarb, Allan Hanbury and Gábor Recski	137

<i>Comparing Grammatical Theories of Code-Mixing</i> Adithya Pratapa and Monojit Choudhury	158
<i>Improving Punctuation Restoration for Speech Transcripts via External Data</i> Xue-Yong Fu, Cheng Chen, Md Tahmid Rahman Laskar, Shashi Bhushan and Simon Corston-Oliver	168
<i>Learning to Rank Question Answer Pairs with Bilateral Contrastive Data Augmentation</i> Yang Deng, Wenxuan Zhang and Wai Lam	175
<i>Mitigation of Diachronic Bias in Fake News Detection Dataset</i> Taichi Murayama, Shoko Wakamiya and Eiji ARAMAKI	182
<i>Understanding the Impact of UGC Specificities on Translation Quality</i> José Carlos Rosales Núñez, Djamé Seddah and Guillaume Wisniewski	189
<i>Noisy UGC Translation at the Character Level: Revisiting Open-Vocabulary Capabilities and Robustness of Char-Based Models</i> José Carlos Rosales Núñez, Guillaume Wisniewski and Djamé Seddah	199
<i>Changes in Twitter geolocations: Insights and suggestions for future usage</i> Anna Kruspe, Matthias Häberle, Eike J. Hoffmann, Samyo Rode-Hasinger, Karam Abdulahhad and Xiao Xiang Zhu	212
<i>ConQuest: Contextual Question Paraphrasing through Answer-Aware Synthetic Question Generation</i> Mostafa Mirshekari, Jing Gu and Aaron Sisto	222
<i>NADE: A Benchmark for Robust Adverse Drug Events Extraction in Face of Negations</i> Simone Scaboro, Beatrice Portelli, Emmanuele Chersoni, Enrico Santus and Giuseppe Serra ..	230
<i>SpanAlign: Efficient Sequence Tagging Annotation Projection into Translated Data applied to Cross-Lingual Opinion Mining</i> Léo Jacqmin, Gabriel Marzinotto, Justyna Gromada, Ewelina Szczekocka, Robert Kołodyński and Géraldine Damnati	238
<i>A Novel Framework for Detecting Important Subevents from Crisis Events via Dynamic Semantic Graphs</i> Evangelia Spiliopoulou, Tanay Kumar Saha, Joel Tetreault and Alejandro Jaimes	249
<i>Synthetic Data Generation and Multi-Task Learning for Extracting Temporal Information from Health-Related Narrative Text</i> Heereen Shim, Dietwig Lowet, Stijn Luca and Bart Vanrumste	260
<i>Neural-based RST Parsing And Analysis In Persuasive Discourse</i> Jinfen Li and Lu Xiao	274
<i>BART for Post-Correction of OCR Newspaper Text</i> Elizabeth Soper, Stanley Fujimoto and Yen-Yun Yu	284
<i>Coping with Noisy Training Data Labels in Paraphrase Detection</i> Teemu Vahtola, Mathias Creutz, Eetu Sjöblom and Sami Itkonen	291
<i>Knowledge Distillation with Noisy Labels for Natural Language Understanding</i> Shivendra Bhardwaj, Abbas Ghaddar, Ahmad Rashid, Khalil Bibi, Chengyang Li, Ali Ghodsi, Phillippe Langlais and Mehdi Rezagholizadeh	297

<i>Integrating Transformers and Knowledge Graphs for Twitter Stance Detection</i> Thomas Clark, Costanza Conforti, Fangyu Liu, Zaiqiao Meng, Ehsan Shareghi and Nigel Collier	304
<i>Detecting Cross-Geographic Biases in Toxicity Modeling on Social Media</i> Sayan Ghosh, Dylan Baker, David Jurgens and Vinodkumar Prabhakaran	313
<i>Detection of Puffery on the English Wikipedia</i> Amanda Bertsch and Steven Bethard	329
<i>Robustness and Sensitivity of BERT Models Predicting Alzheimer’s Disease from Text</i> Jekaterina Novikova	334
<i>Understanding Model Robustness to User-generated Noisy Texts</i> Jakub Náplava, Martin Popel, Milan Straka and Jana Straková	340
<i>CIDEr-R: Robust Consensus-based Image Description Evaluation</i> Gabriel Oliveira dos Santos, Esther Luna Colombini and Sandra Avila	351
<i>Improved Named Entity Recognition for Noisy Call Center Transcripts</i> Sam Davidson, Jordan Hosier, Yu Zhou and Vijay Gurbani	361
<i>Contrapositive Local Class Inference</i> Omid Kashefi and Rebecca Hwa	371
<i>Improved Multilingual Language Model Pretraining for Social Media Text via Translation Pair Prediction</i> Shubhanshu Mishra and Aria Haghighi	381
<i>Co-training for Commit Classification</i> Jian Yi David Lee and Hai Leong Chieu	389
<i>Study of Manifestation of Civil Unrest on Twitter</i> Abhinav Chinta, Jingyu Zhang, Alexandra DeLucia, Mark Dredze and Anna L. Buczak	396
<i>The Korean Morphologically Tight-Fitting Tokenizer for Noisy User-Generated Texts</i> Sangah Lee and Hyopil Shin	410
<i>Character Transformations for Non-Autoregressive GEC Tagging</i> Milan Straka, Jakub Náplava and Jana Straková	417
<i>Can Character-based Language Models Improve Downstream Task Performances In Low-Resource And Noisy Language Scenarios?</i> Arij riabi, Benoît Sagot and Djamé Seddah	423
<i>"Something Something Hota Hai!" An Explainable Approach towards Sentiment Analysis on Indian Code-Mixed Data</i> Aman Priyanshu, Aleti Vardhan, Sudarshan Sivakumar, Supriti Vijay and Nipuna Chhabra	437
<i>BERTweetFR : Domain Adaptation of Pre-Trained Language Models for French Tweets</i> Yanzhu Guo, Virgile Rennard, Christos Xypolopoulos and Michalis Vazirgiannis	445
<i>To What Extent Does Lexical Normalization Help English-as-a-Second Language Learners to Read Noisy English Texts?</i> Yo Ehara	451

<i>Multilingual Sequence Labeling Approach to solve Lexical Normalization</i>	
Divesh Kubal and Apurva Nagvenkar	457
<i>Sesame Street to Mount Sinai: BERT-constrained character-level Moses models for multilingual lexical normalization</i>	
Yves Scherrer and Nikola Ljubešić	465
<i>Sequence-to-Sequence Lexical Normalization with Multilingual Transformers</i>	
Ana-Maria Bucur, Adrian Cosma and Liviu P. Dinu	473
<i>ÚFAL at MultiLexNorm 2021: Improving Multilingual Lexical Normalization by Fine-tuning ByT5</i>	
David Samuel and Milan Straka	483
<i>MultiLexNorm: A Shared Task on Multilingual Lexical Normalization</i>	
Rob van der Goot, Alan Ramponi, Arkaitz Zubiaga, Barbara Plank, Benjamin Muller, Iñaki San Vicente Roncal, Nikola Ljubešić, Özlem Çetinoğlu, Rahmad Mahendra, Talha Çolakoğlu, Timothy Baldwin, Tommaso Caselli and Wladimir Sidorenko	493
<i>CL-MoNoise: Cross-lingual Lexical Normalization</i>	
Rob van der Goot	510

Conference Program

44511

12:00–12:00 Session WNUT: WNUT

Text Simplification for Comprehension-based Question-Answering

Tanvi Dadu, Kartikey Pant, Seema Nagar, Ferdous Barbhuiya and Kuntal Dey

Finding the needle in a haystack: Extraction of Informative COVID-19 Danish Tweets

Benjamin Olsen and Barbara Plank

Detecting Depression in Thai Blog Posts: a Dataset and a Baseline

Mika Hämmäläinen, Pattama Patpong, Khalid Alnajjar, Niko Partanen and Jack Rueter

Keyphrase Extraction with Incomplete Annotated Training Data

Yanfei Lei, Chunming Hu, Guanghui Ma and Richong Zhang

Fine-grained Temporal Relation Extraction with Ordered-Neuron LSTM and Graph Convolutional Networks

Minh Tran Phu, Minh Van Nguyen and Thien Huu Nguyen

Does It Happen? Multi-hop Path Structures for Event Factuality Prediction with Graph Transformer Networks

Duong Le and Thien Huu Nguyen

Google-trickers, Yaminjeongeum, and Leetspeak: An Empirical Taxonomy for Intentionally Noisy User-Generated Text

Won Ik Cho and Soomin Kim

Description-based Label Attention Classifier for Explainable ICD-9 Classification

Malte Feucht, Zhiliang Wu, Sophia Althammer and Volker Tresp

A Text Editing Approach to Joint Japanese Word Segmentation, POS Tagging, and Lexical Normalization

Shohei Higashiyama, Masao Utiyama, Taro Watanabe and Eiichiro Sumita

Intrinsic evaluation of language models for code-switching

Sik Feng Cheong, Hai Leong Chieu and Jing Lim

44511 (continued)

Can images help recognize entities? A study of the role of images for Multimodal NER

Shuguang Chen, Gustavo Aguilar, Leonardo Neves and Thamar Solorio

Perceived and Intended Sarcasm Detection with Graph Attention Networks

Joan Plepi and Lucie Flek

Hierarchical Character Tagger for Short Text Spelling Error Correction

Mengyi Gao, Canran Xu and Peng Shi

Common Sense Bias in Semantic Role Labeling

Heather Lent and Anders Søgaard

PoliWAM: An Exploration of a Large Scale Corpus of Political Discussions on WhatsApp Messenger

Vivek Srivastava and Mayank Singh

ParsTwiNER: A Corpus for Named Entity Recognition at Informal Persian

MohammadMahdi Aghajani, AliAkbar Badri and Hamid Beigy

DreamDrug - A crowdsourced NER dataset for detecting drugs in darknet markets

Johannes Bogensperger, Sven Schlarb, Allan Hanbury and Gábor Recski

Comparing Grammatical Theories of Code-Mixing

Adithya Pratapa and Monojit Choudhury

Improving Punctuation Restoration for Speech Transcripts via External Data

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

Learning to Rank Question Answer Pairs with Bilateral Contrastive Data Augmentation

Yang Deng, Wenxuan Zhang and Wai Lam

Mitigation of Diachronic Bias in Fake News Detection Dataset

Taichi Murayama, Shoko Wakamiya and Eiji ARAMAKI

Understanding the Impact of UGC Specificities on Translation Quality

José Carlos Rosales Núñez, Djamé Seddah and Guillaume Wisniewski

44511 (continued)

Noisy UGC Translation at the Character Level: Revisiting Open-Vocabulary Capabilities and Robustness of Char-Based Models

José Carlos Rosales Núñez, Guillaume Wisniewski and Djamé Seddah

Changes in Twitter geolocations: Insights and suggestions for future usage

Anna Kruspe, Matthias Häberle, Eike J. Hoffmann, Samyo Rode-Hasinger, Karam Abdulahhad and Xiao Xiang Zhu

ConQuest: Contextual Question Paraphrasing through Answer-Aware Synthetic Question Generation

Mostafa Mirshekari, Jing Gu and Aaron Sisto

NADE: A Benchmark for Robust Adverse Drug Events Extraction in Face of Negations

Simone Scabro, Beatrice Portelli, Emmanuele Chersoni, Enrico Santus and Giuseppe Serra

SpanAlign: Efficient Sequence Tagging Annotation Projection into Translated Data applied to Cross-Lingual Opinion Mining

Léo Jacqmin, Gabriel Marzinotto, Justyna Gromada, Ewelina Szczekocka, Robert Kołodyński and Géraldine Damnati

A Novel Framework for Detecting Important Subevents from Crisis Events via Dynamic Semantic Graphs

Evangelia Spiliopoulou, Tanay Kumar Saha, Joel Tetreault and Alejandro Jaimes

Synthetic Data Generation and Multi-Task Learning for Extracting Temporal Information from Health-Related Narrative Text

Heereen Shim, Dietwig Lowet, Stijn Luca and Bart Vanrumste

Neural-based RST Parsing And Analysis In Persuasive Discourse

Jinfen Li and Lu Xiao

BART for Post-Correction of OCR Newspaper Text

Elizabeth Soper, Stanley Fujimoto and Yen-Yun Yu

Coping with Noisy Training Data Labels in Paraphrase Detection

Teemu Vahtola, Mathias Creutz, Eetu Sjöblom and Sami Itkonen

Knowledge Distillation with Noisy Labels for Natural Language Understanding

Shivendra Bhardwaj, Abbas Ghaddar, Ahmad Rashid, Khalil Bibi, Chengyang Li, Ali Ghodsi, Phillippe Langlais and Mehdi Rezagholizadeh

Integrating Transformers and Knowledge Graphs for Twitter Stance Detection

Thomas Clark, Costanza Conforti, Fangyu Liu, Zaiqiao Meng, Ehsan Shareghi and Nigel Collier

44511 (continued)

Detecting Cross-Geographic Biases in Toxicity Modeling on Social Media

Sayan Ghosh, Dylan Baker, David Jurgens and Vinodkumar Prabhakaran

Detection of Puffery on the English Wikipedia

Amanda Bertsch and Steven Bethard

Robustness and Sensitivity of BERT Models Predicting Alzheimer's Disease from Text

Jekaterina Novikova

Understanding Model Robustness to User-generated Noisy Texts

Jakub Náplava, Martin Popel, Milan Straka and Jana Straková

CIDEr-R: Robust Consensus-based Image Description Evaluation

Gabriel Oliveira dos Santos, Esther Luna Colombini and Sandra Avila

Improved Named Entity Recognition for Noisy Call Center Transcripts

Sam Davidson, Jordan Hosier, Yu Zhou and Vijay Gurbani

Contrapositive Local Class Inference

Omid Kashefi and Rebecca Hwa

Improved Multilingual Language Model Pretraining for Social Media Text via Translation Pair Prediction

Shubhanshu Mishra and Aria Haghighi

Co-training for Commit Classification

Jian Yi David Lee and Hai Leong Chieu

Study of Manifestation of Civil Unrest on Twitter

Abhinav Chinta, Jingyu Zhang, Alexandra DeLucia, Mark Dredze and Anna L. Buczak

The Korean Morphologically Tight-Fitting Tokenizer for Noisy User-Generated Texts

Sangah Lee and Hyopil Shin

Character Transformations for Non-Autoregressive GEC Tagging

Milan Straka, Jakub Náplava and Jana Straková

44511 (continued)

Can Character-based Language Models Improve Downstream Task Performances In Low-Resource And Noisy Language Scenarios?

Arij riabi, Benoît Sagot and Djamé Seddah

"Something Something Hota Hai!" An Explainable Approach towards Sentiment Analysis on Indian Code-Mixed Data

Aman Priyanshu, Aleti Vardhan, Sudarshan Sivakumar, Supriti Vijay and Nipuna Chhabra

BERTweetFR : Domain Adaptation of Pre-Trained Language Models for French Tweets

Yanzhu Guo, Virgile Rennard, Christos Xypolopoulos and Michalis Vazirgiannis

To What Extent Does Lexical Normalization Help English-as-a-Second Language Learners to Read Noisy English Texts?

Yo Ehara

Multilingual Sequence Labeling Approach to solve Lexical Normalization

Divesh Kubal and Apurva Nagvenkar

Sesame Street to Mount Sinai: BERT-constrained character-level Moses models for multilingual lexical normalization

Yves Scherrer and Nikola Ljubešić

Sequence-to-Sequence Lexical Normalization with Multilingual Transformers

Ana-Maria Bucur, Adrian Cosma and Liviu P. Dinu

ÚFAL at MultiLexNorm 2021: Improving Multilingual Lexical Normalization by Fine-tuning ByT5

David Samuel and Milan Straka

MultiLexNorm: A Shared Task on Multilingual Lexical Normalization

Rob van der Goot, Alan Ramponi, Arkaitz Zubiaga, Barbara Plank, Benjamin Muller, Iñaki San Vicente Roncal, Nikola Ljubešić, Özlem Çetinoğlu, Rahmad Mahendra, Talha Çolakoğlu, Timothy Baldwin, Tommaso Caselli and Wladimir Sidorenko

CL-MoNoise: Cross-lingual Lexical Normalization

Rob van der Goot

