

# **2nd Workshop on Noisy User-generated Text (WNUT 2016)**

Osaka, Japan  
11 December 2016

ISBN: 978-1-5108-3319-7

**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© (2016) by the Association for Computational Linguistics  
All rights reserved.

Printed by Curran Associates, Inc. (2016)

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](mailto: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](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

## Table of Contents

<i>Processing non-canonical or noisy text: fortuitous data to the rescue</i> Barbara Plank .....	1
<i>From Entity Linking to Question Answering – Recent Progress on Semantic Grounding Tasks</i> Ming-Wei Chang .....	2
<i>DISAANA and D-SUMM: Large-scale Real Time NLP Systems for Analyzing Disaster Related Reports in Tweets</i> Kentaro Torisawa .....	3
<i>Private or Corporate? Predicting User Types on Twitter</i> Nikola Ljubešić and Darja Fišer .....	4
<i>From Noisy Questions to Minecraft Texts: Annotation Challenges in Extreme Syntax Scenario</i> Héctor Martínez Alonso, Djamé Seddah and Benoît Sagot .....	13
<i>Disaster Analysis using User-Generated Weather Report</i> Yasunobu Asakura, Masatsugu Hangyo and Mamoru Komachi .....	24
<i>Veracity Computing from Lexical Cues and Perceived Certainty Trends</i> Uwe Reichel and Piroska Lendvai .....	33
<i>A Simple but Effective Approach to Improve Arabizi-to-English Statistical Machine Translation</i> Marlies van der Wees, Arianna Bisazza and Christof Monz .....	43
<i>Name Variation in Community Question Answering Systems</i> Anietie Andy, Satoshi Sekine, Mugizi Rwebangira and Mark Dredze .....	51
<i>Whose Nickname is This? Recognizing Politicians from Their Aliases</i> Wei-Chung Wang, Hung-Chen Chen, Zhi-Kai Ji, Hui-I Hsiao, Yu-Shian Chiu and Lun-Wei Ku ..	61
<i>Towards Accurate Event Detection in Social Media: A Weakly Supervised Approach for Learning Implicit Event Indicators</i> Ajit Jain, Girish Kasiviswanathan and Ruihong Huang .....	70
<i>Unsupervised Stemmer for Arabic Tweets</i> Fahad Albogamy and Allan Ramsay .....	78
<i>Topic Stability over Noisy Sources</i> Jing Su, Derek Greene and Oisín Boydell .....	85
<i>Analysis of Twitter Data for Postmarketing Surveillance in Pharmacovigilance</i> Julie Pain, Jessie Levacher, Adam Quinquenel and Anja Belz .....	94
<i>Named Entity Recognition and Hashtag Decomposition to Improve the Classification of Tweets</i> Billal Belainine, Alessandro Fonseca and Fatiha Sadat .....	102
<i>Exploring Word Embeddings for Unsupervised Textual User-Generated Content Normalization</i> Thales Felipe Costa Bertaglia and Maria das Graças Volpe Nunes .....	112
<i>How Document Pre-processing affects Keyphrase Extraction Performance</i> Florian Boudin, Hugo Mougard and Damien Cram .....	121

<i>Japanese Text Normalization with Encoder-Decoder Model</i>	
Taishi Ikeda, Hiroyuki Shindo and Yuji Matsumoto .....	129
<i>Results of the WNUT16 Named Entity Recognition Shared Task</i>	
Benjamin Strauss, Bethany Toma, Alan Ritter, Marie-Catherine de Marneffe and Wei Xu .....	138
<i>Bidirectional LSTM for Named Entity Recognition in Twitter Messages</i>	
Nut Limsopatham and Nigel Collier .....	145
<i>Learning to recognise named entities in tweets by exploiting weakly labelled data</i>	
Kurt Junshean Espinosa, Riza Theresa Batista-Navarro and Sophia Ananiadou .....	153
<i>Feature-Rich Twitter Named Entity Recognition and Classification</i>	
Utpal Kumar Sikdar and Björn Gambäck .....	164
<i>Learning to Search for Recognizing Named Entities in Twitter</i>	
Ioannis Partalas, Cédric Lopez, Nadia Derbas and Ruslan Kalitvianski .....	171
<i>DeepNNER: Applying BLSTM-CNNs and Extended Lexicons to Named Entity Recognition in Tweets</i>	
Fabrice Dugas and Eric Nichols .....	178
<i>ASU: An Experimental Study on Applying Deep Learning in Twitter Named Entity Recognition.</i>	
Michel Naim Gerguis, Cherif Salama and M. Watheq El-Kharashi .....	188
<i>UQAM-NTL: Named entity recognition in Twitter messages</i>	
Ngoc Tan LE, Fatma Mallek and Fatiha Sadat .....	197
<i>Semi-supervised Named Entity Recognition in noisy-text</i>	
Shubhanshu Mishra and Jana Diesner .....	203
<i>Twitter Geolocation Prediction Shared Task of the 2016 Workshop on Noisy User-generated Text</i>	
Bo Han, Afshin Rahimi, Leon Derczynski and Timothy Baldwin .....	213
<i>CSIRO Data61 at the WNUT Geo Shared Task</i>	
Gaya Jayasinghe, Brian Jin, James Mchugh, Bella Robinson and Stephen Wan .....	218
<i>Geolocation Prediction in Twitter Using Location Indicative Words and Textual Features</i>	
Lianhua Chi, Kwan Hui Lim, Nebula Alam and Christopher J. Butler .....	227
<i>A Simple Scalable Neural Networks based Model for Geolocation Prediction in Twitter</i>	
Yasuhide Miura, Motoki Taniguchi, Tomoki Taniguchi and Tomoko Ohkuma .....	235