

3rd Workshop on Noisy User-generated Text (W-NUT 2017)

Copenhagen, Denmark
7 September 2017

ISBN: 978-1-5108-4753-8

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© (2017) by the Association for Computational Linguistics
All rights reserved.

Printed by Curran Associates, Inc. (2017)

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>Boundary-based MWE segmentation with text partitioning</i> Jake Williams	1
<i>Towards the Understanding of Gaming Audiences by Modeling Twitch Emotes</i> Francesco Barbieri, Luis Espinosa Anke, Miguel Ballesteros, Juan Soler and Horacio Saggion ..	11
<i>Churn Identification in Microblogs using Convolutional Neural Networks with Structured Logical Knowledge</i> Mourad Gridach, Hatem Haddad and Hala Mulki	21
<i>To normalize, or not to normalize: The impact of normalization on Part-of-Speech tagging</i> Rob van der Goot, Barbara Plank and Malvina Nissim	31
<i>Constructing an Alias List for Named Entities during an Event</i> Anietie Andy, Mark Dredze, Mugizi Rwebangira and Chris Callison-Burch	40
<i>Incorporating Metadata into Content-Based User Embeddings</i> Linzi Xing and Michael J. Paul	45
<i>Simple Queries as Distant Labels for Predicting Gender on Twitter</i> Chris Emmerly, Grzegorz Chrupala and Walter Daelemans	50
<i>A Dataset and Classifier for Recognizing Social Media English</i> Su Lin Blodgett, Johnny Wei and Brendan O'Connor	56
<i>Evaluating hypotheses in geolocation on a very large sample of Twitter</i> Bahar Salehi and Anders Søgaard	62
<i>The Effect of Error Rate in Artificially Generated Data for Automatic Preposition and Determiner Correction</i> Fraser Bowen, Jon Dehdari and Josef Van Genabith	68
<i>An Entity Resolution Approach to Isolate Instances of Human Trafficking Online</i> Chirag Nagpal, Kyle Miller, Benedikt Boecking and Artur Dubrawski	77
<i>Noisy Uyghur Text Normalization</i> Osman Tursun and Ruket Cakici	85
<i>Crowdsourcing Multiple Choice Science Questions</i> Johannes Welbl, Nelson F. Liu and Matt Gardner	94
<i>A Text Normalisation System for Non-Standard English Words</i> Emma Flint, Elliot Ford, Olivia Thomas, Andrew Caines and Paula Buttery	107
<i>Huntsville, hospitals, and hockey teams: Names can reveal your location</i> Bahar Salehi, Dirk Hovy, Eduard Hovy and Anders Søgaard	116
<i>Improving Document Clustering by Removing Unnatural Language</i> Myungha Jang, Jinho D. Choi and James Allan	122
<i>Lithium NLP: A System for Rich Information Extraction from Noisy User Generated Text on Social Media</i> Preeti Bhargava, Nemanja Spasojevic and Guoning Hu	131

<i>Results of the WNUT2017 Shared Task on Novel and Emerging Entity Recognition</i>	
Leon Derczynski, Eric Nichols, Marieke van Erp and Nut Limsopatham	140
<i>A Multi-task Approach for Named Entity Recognition in Social Media Data</i>	
Gustavo Aguilar, Suraj Maharjan, Adrian Pastor López Monroy and Thamar Solorio	148
<i>Distributed Representation, LDA Topic Modelling and Deep Learning for Emerging Named Entity Recognition from Social Media</i>	
Patrick Jansson and Shuhua Liu	154
<i>Multi-channel BiLSTM-CRF Model for Emerging Named Entity Recognition in Social Media</i>	
Bill Y. Lin, Frank Xu, Zhiyi Luo and Kenny Zhu	160
<i>Transfer Learning and Sentence Level Features for Named Entity Recognition on Tweets</i>	
Pius von Däniken and Mark Cieliebak	166
<i>Context-Sensitive Recognition for Emerging and Rare Entities</i>	
Jake Williams and Giovanni Santia	172
<i>A Feature-based Ensemble Approach to Recognition of Emerging and Rare Named Entities</i>	
Utpal Kumar Sikdar and Björn Gambäck	177