WNUT 2016

The 2nd Workshop on Noisy User-generated Text

Proceedings of the Workshop

December 11, 2016
Osaka, Japan

ISBN978-4-87974-707-5

Preface

This volume contains papers from the 2nd Workshop on on Noisy User-generated Text (W-NUT)

Organisers

Bo Han (Hugo AI)

Alan Ritter (The Ohio State University)

Leon Derczynski (The University of Sheffield)

Wei Xu (The Ohio State University)

Tim Baldwin (The University of Melbourne)

Programme Committee

David Bamman (University of California, Berkeley)

Kalina Bontcheva (University of Sheffield)

Claire Cardie (Cornell University)

Colin Cherry (National Research Council Canada)

Grzegorz Chrupała (Tilburg University)

Marina Danilevsky (IBM Research)

Seza Doğruöz (Tilburg University)

Heba Elfardy (Columbia University)

Noura Farra (Columbia University)

Eric Fosler-Lussier (The Ohio State University)

Dan Garrette (University of Washington)

Kevin Gimpel (Toyota Technological Institute at Chicago)

Weiwei Guo (Yahoo! Research)

Ben Hachey (Hugo AI)

Masato Hagiwara (Duolingo)

Hua He (University of Maryland)

Ed Hovy (Carnegie Mellon University)

Jing Jiang (Singapore Management University)

Anna Jørgensen (University of Amsterdam)

Nobuhiro Kaji (Yahoo! Research)

Emre Kiciman (Microsoft Research)

Chen Li (University of Texas at Dallas)

Junyi Jessy Li (University of Pennsylvania)

Wang Ling (Google DeepMind)

Fei Liu (University of Central Florida)

Huan Liu (Arizona State University)

Héctor Martínez Alonso (INRIA/University Paris Diderot)

Rada Mihalcea (University of Michigan)

Smaranda Muresan (Columbia University)

Preslav Nakov (Qatar Computing Research Institute)

Naoaki Okazaki (Tohoku University)

Miles Osborne (Bloomberg)

Ellie Pavlick (University of Pennsylvania)

Daniel Preoţiuc-Pietro (University of Pennsylvania)

Will Radford (Hugo AI)

Afshin Rahimi (The University of Melbourne)

Shourya Roy (Xerox Research)

Alla Rozovskaya (City University of New York)

Derek Ruths (McGill University)

Andrew Schwartz (Stony Brook University)

Djamé Seddah (University Paris-Sorbonne)

Richard Sproat (Google Research)

Anders Søgaard (University of Copenhagen)

Benjamin Strauss (The Ohio State University)

Jeniya Tabassum (The Ohio State University)

Joel Tetreault (Yahoo! Research)

Marlies van der Wees (University of Amsterdam)

Svitlana Volkova (Pacific Northwest National Laboratory)

Byron C. Wallace (University of Texas at Austin)

Xiaojun Wan (Peking University)

Jun-Ming Xu (University of Wisconsin-Madison)

Diyi Yang (Carnegie Mellon University)

Yi Yang (Georgia Tech)

Guido Zarrella (MITRE)

Ming Zhou (Microsoft Research)

Table of Contents

Workshop Program

December 11, 2016

9.00 **Opening**

Invited talk

9.10 *DISAANA and D-SUMM: Large-scale Real Time NLP Systems for Analyzing Disaster Related Reports in Tweets*
Kentaro Torisawa

Research talks

9.55 *Private or Corporate? Predicting User Types on Twitter*
Nikola Ljubešić and Darja Fišer

10.05 *From Noisy Questions to Minecraft Texts: Annotation Challenges in Extreme Syntax Scenario*
Héctor Martínez Alonso, Djamé Seddah and Benoît Sagot

10.15 *Disaster Analysis using User-Generated Weather Report*
Yasunobu Asakura, Masatsugu Hangyo and Mamoru Komachi

10.25 *Veracity Computing from Lexical Cues and Perceived Certainty Trends*
Uwe Reichel and Piroska Lendvai

10.35 *Exploring Word Embeddings for Unsupervised Textual User-Generated Content Normalization*
Thales Felipe Costa Bertaglia and Maria das Graças Volpe Nunes

10.45 *Name Variation in Community Question Answering Systems*
Anietie Andy, Satoshi Sekine, Mugizi Rwebangira and Mark Dredze

10.55 **Research posters**

Whose Nickname is This? Recognizing Politicians from Their Aliases
Wei-Chung Wang, Hung-Chen Chen, Zhi-Kai Ji, Hui-I Hsiao, Yu-Shian Chiu and Lun-Wei Ku

Towards Accurate Event Detection in Social Media: A Weakly Supervised Approach for Learning Implicit Event Indicators
Ajit Jain, Girish Kasiviswanathan and Ruihong Huang

Unsupervised Stemmer for Arabic Tweets
Fahad Albogamy and Allan Ramsay

Topic Stability over Noisy Sources
Jing Su, Derek Greene and Oisin Boydell

Analysis of Twitter Data for Postmarketing Surveillance in Pharmacovigilance
Julie Pain, Jessie Levacher, Adam Quinquenel and Anja Belz

Named Entity Recognition and Hashtag Decomposition to Improve the Classification of Tweets
Billal Belainine, Alexsandro Fonseca and Fatiha Sadat

A Simple but Effective Approach to Improve Arabizi-to-English Statistical Machine Translation
Marlies van der Wees, Arianna Bisazza and Christof Monz

How Document Pre-processing affects Keyphrase Extraction Performance
Florian Boudin, Hugo Mougard and Damien Cram

Japanese Text Normalization with Encoder-Decoder Model
Taishi Ikeda, Hiroyuki Shindo and Yuji Matsumoto

December 11, 2016 (continued)

Invited talk

14.00 *From Entity Linking to Question Answering – Recent Progress on Semantic Grounding Tasks*
Ming-Wei Chang

14.45 **Shared task papers**

Results of the WNUT16 Named Entity Recognition Shared Task
Benjamin Strauss, Bethany Toma, Alan Ritter, Marie-Catherine de Marneffe and Wei Xu

Bidirectional LSTM for Named Entity Recognition in Twitter Messages
Nut Limsopatham and Nigel Collier

Learning to recognise named entities in tweets by exploiting weakly labelled data
Kurt Junshean Espinosa, Riza Theresa Batista-Navarro and Sophia Ananiadou

Feature-Rich Twitter Named Entity Recognition and Classification
Utpal Kumar Sikdar and Björn Gambäck

Learning to Search for Recognizing Named Entities in Twitter
Ioannis Partalas, Cédric Lopez, Nadia Derbas and Ruslan Kalitvianski

DeepNNNER: Applying BLSTM-CNNs and Extended Lexicons to Named Entity Recognition in Tweets
Fabrice Dugas and Eric Nichols

ASU: An Experimental Study on Applying Deep Learning in Twitter Named Entity Recognition.
Michel Naim Gerguis, Cherif Salama and M. Watheq El-Kharashi

UQAM-NTL: Named entity recognition in Twitter messages
Ngoc Tan LE, Fatma Mallek and Fatiha Sadat

Semi-supervised Named Entity Recognition in noisy-text
Shubhanshu Mishra and Jana Diesner

Twitter Geolocation Prediction Shared Task of the 2016 Workshop on Noisy User-generated Text
Bo Han, Afshin Rahimi, Leon Derczynski and Timothy Baldwin

CSIRO Data61 at the WNUT Geo Shared Task
Gaya Jayasinghe, Brian Jin, James Mchugh, Bella Robinson and Stephen Wan

December 11, 2016 (continued)

Geolocation Prediction in Twitter Using Location Indicative Words and Textual Features
Lianhua Chi, Kwan Hui Lim, Nebula Alam and Christopher J. Butler

A Simple Scalable Neural Networks based Model for Geolocation Prediction in Twitter
Yasuhide Miura, Motoki Taniguchi, Tomoki Taniguchi and Tomoko Ohkuma

Invited talk

16.10 *Processing non-canonical or noisy text: fortuitous data to the rescue*
Barbara Plank