Workshop on Noisy User-generated Text 2015

Co-located with ACL-IJCNLP 2015

Beijing, China
31 July 2015

ISBN: 978-1-5108-0953-6

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© (2015) by the Association for Computational Linguistics
All rights reserved.

Printed by Curran Associates, Inc. (2015)

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: 845-758-2634
Email: curran@proceedings.com
Web: www.proceedings.com

Table of Contents

Minority Language Twitter: Part-of-Speech Tagging and Analysis of Irish Tweets
Teresa Lynn, Kevin Scannell and Eimear Maguire 1

Challenges of studying and processing dialects in social media
Anna Jørgensen, Dirk Hovy and Anders Søgaard 9

Toward Tweets Normalization Using Maximum Entropy
Mohammad Arshi Saloot, Norisma Idris, Liyana Shuib, Ram Gopal Raj and AiTi Aw 19

Five Shades of Noise: Analyzing Machine Translation Errors in User-Generated Text
Marlies van der Wees, Arianna Bisazza and Christof Monz 28

A Normalizer for UGC in Brazilian Portuguese
Magali Sanches Duran, Maria das Graças Volpe Nunes and Lucas Avanço 38

USFD: Twitter NER with Drift Compensation and Linked Data
Leon Derczynski, Isabelle Augenstein and Kalina Bontcheva 48

NRC: Infused Phrase Vectors for Named Entity Recognition in Twitter
Colin Cherry, Hongyu Guo and Chengbi Dai 54

IITP: Multiobjective Differential Evolution based Twitter Named Entity Recognition
Md Shad Akhtar, Utpal Kumar Sikdar and Asif Ekbal 61

Data Adaptation for Named Entity Recognition on Tweets with Features-Rich CRF
Tian Tian, Marco Dinarelli and Isabelle Tellier 68

Hallym: Named Entity Recognition on Twitter with Word Representation
Eun-Suk Yang and Yu-Seop Kim 72

IHS_RD: Lexical Normalization for English Tweets
Dmitry Supranovich and Viachaslau Patsepnia 78

Bekli:A Simple Approach to Twitter Text Normalization.
Russell Beckley 82

NCSU-SAS-Ning: Candidate Generation and Feature Engineering for Supervised Lexical Normalization
Ning Jin 87

DCU-ADAPT: Learning Edit Operations for Microblog Normalisation with the Generalised Perceptron
Joachim Wagner and Jennifer Foster 93

LYSGROUP: Adapting a Spanish microtext normalization system to English.
Yerai Doval Mosquera, Jesús Vilares and Carlos Gómez-Rodríguez 99

IITP: Hybrid Approach for Text Normalization in Twitter
Md Shad Akhtar, Utpal Kumar Sikdar and Asif Ekbal 106

NCSU_SAS_WOOKHEE: A Deep Contextual Long-Short Term Memory Model for Text Normalization
Wookhee Min and Bradford Mott 111

USZEGED: Correction Type-sensitive Normalization of English Tweets Using Efficiently Indexed n-gram Statistics
Gábor Berend and Ervin Tasnádi 120

Shared Tasks of the 2015 Workshop on Noisy User-generated Text: Twitter Lexical Normalization and Named Entity Recognition
Timothy Baldwin, Marie-Catherine de Marneffe, Bo Han, Young-Bum Kim, Alan Ritter and Wei Xu 126

Enhancing Named Entity Recognition in Twitter Messages Using Entity Linking
Ikuya Yamada, Hideaki Takeda and Yoshiyasu Takefuji 136

Improving Twitter Named Entity Recognition using Word Representations
Zhiqiang Toh, Bin Chen and Jian Su 141

Multimedia Lab @ ACL WNUT NER Shared Task: Named Entity Recognition for Twitter Microposts using Distributed Word Representations
Fréderic Godin, Baptist Vandersmissen, Wesley De Neve and Rik Van de Walle 146

NCSU_SAS_SAM: Deep Encoding and Reconstruction for Normalization of Noisy Text
Samuel Leeman-Munk, James Lester and James Cox 154