

# **First Workshop on Abusive Language Online 2017**

Held at ACL 2017

Vancouver, Canada  
4 August 2017

ISBN: 978-1-5108-4580-0

**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](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>Dimensions of Abusive Language on Twitter</i> Isabelle Clarke and Dr. Jack Grieve .....	1
<i>Constructive Language in News Comments</i> Varada Kolhatkar and Maite Taboada .....	11
<i>Rephrasing Profanity in Chinese Text</i> Hui-Po Su, Zhen-Jie Huang, Hao-Tsung Chang and Chuan-Jie Lin .....	18
<i>Deep Learning for User Comment Moderation</i> John Pavlopoulos, Prodromos Malakasiotis and Ion Androutsopoulos .....	25
<i>Class-based Prediction Errors to Detect Hate Speech with Out-of-vocabulary Words</i> Joan Serrà, Ilias Leontiadis, Dimitris Spathis, Gianluca Stringhini, Jeremy Blackburn and Athena Vakali .....	36
<i>One-step and Two-step Classification for Abusive Language Detection on Twitter</i> Ji Ho Park and Pascale Fung .....	41
<i>Legal Framework, Dataset and Annotation Schema for Socially Unacceptable Online Discourse Practices in Slovene</i> Darja Fišer, Tomaž Erjavec and Nikola Ljubešić .....	46
<i>Abusive Language Detection on Arabic Social Media</i> Hamdy Mubarak, Kareem Darwish and Walid Magdy .....	52
<i>Vectors for Counterspeech on Twitter</i> Lucas Wright, Derek Ruths, Kelly P Dillon, Haji Mohammad Saleem and Susan Benesch .....	57
<i>Detecting Nastiness in Social Media</i> Niloofar Safi Samghabadi, Suraj Maharjan, Alan Sprague, Raquel Diaz-Sprague and Thamar Solorio .....	63
<i>Technology Solutions to Combat Online Harassment</i> George Kennedy, Andrew McCollough, Edward Dixon, Alexei Bastidas, John Ryan, Chris Loo and Saurav Sahay .....	73
<i>Understanding Abuse: A Typology of Abusive Language Detection Subtasks</i> Zeeraq Waseem, Thomas Davidson, Dana Warmusley and Ingmar Weber .....	78
<i>Using Convolutional Neural Networks to Classify Hate-Speech</i> Björn Gambäck and Utpal Kumar Sikdar .....	85
<i>Illegal is not a Noun: Linguistic Form for Detection of Pejorative Nominalizations</i> Alexis Palmer, Melissa Robinson and Kristy K. Phillips .....	91