Third Workshop on Abusive Language Online 2019

Florence, Italy
1 August 2019

ISBN: 978-1-5108-9111-1

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© (2019) by the Association for Computational Linguistics
All rights reserved.

Printed by Curran Associates, Inc. (2019)

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

Subversive Toxicity Detection using Sentiment Information
Eloi Brassard-Gourdeau and Richard Khoury 1

Exploring Deep Multimodal Fusion of Text and Photo for Hate Speech Classification
Fan Yang, Xiaochang Peng, Gargi Ghosh, Reshef Shilon, Hao Ma, Eider Moore and Goran Predovic 11

Detecting harassment in real-time as conversations develop
Wessel Stoop, Florian Kunneman, Antal van den Bosch and Ben Miller 19

Racial Bias in Hate Speech and Abusive Language Detection Datasets
Thomas Davidson, Debasmita Bhattacharya and Ingmar Weber 25

Automated Identification of Verbally Abusive Behaviors in Online Discussions
Srecko Joksimovic, Ryan S. Baker, Jaclyn Ocumpaugh, Juan Miguel L. Andres, Ivan Tot, Elle Yuan Wang and Shane Dawson 36

Multi-label Hate Speech and Abusive Language Detection in Indonesian Twitter
Muhammad Okky Ibrohim and Indra Budi 46

The Discourse of Online Content Moderation: Investigating Polarized User Responses to Changes in Reddit's Quarantine Policy
Qinlan Shen and Carolyn Rose 58

Pay "Attention" to your Context when Classifying Abusive Language
Tuhin Chakrabarty, Kilol Gupta and Smaranda Muresan 70

Challenges and frontiers in abusive content detection
Bertie Vidgen, Alex Harris, Dong Nguyen, Rebekah Tromble, Scott Hale and Helen Margetts . . 80

A Hierarchically-Labeled Portuguese Hate Speech Dataset
Paula Fortuna, João Rocha da Silva, Juan Soler-Company, Leo Wanner and Sérgio Nunes 94

A System to Monitor Cyberbullying based on Message Classification and Social Network Analysis
Stefano Menini, Giovanni Moretti, Michele Corazza, Elena Cabrio, Sara Tonelli and Serena Villata 105

L-HSAB: A Levantine Twitter Dataset for Hate Speech and Abusive Language
Hala Mulki, Hatem Haddad, Chedi Bechikh Ali and Halima Alshabani 111

At the Lower End of Language—Exploring the Vulgar and Obscene Side of German
Elisabeth Eder, Ulrike Krieg-Holz and Udo Hahn 119

Preemptive Toxic Language Detection in Wikipedia Comments Using Thread-Level Context
Mladen Karan and Jan Šnajder 129

Neural Word Decomposition Models for Abusive Language Detection
Sravan Bodapati, Spandana Gella, Kasturi Bhattacharjee and Yaser Al-Onaizan 135

A Platform Agnostic Dual-Strand Hate Speech Detector
Johannes Skjeggestad Meyer and Björn Gambäck 146

Detecting Aggression and Toxicity using a Multi Dimension Capsule Network
Saurabh Srivastava and Prerna Khurana 157

An Impossible Dialogue! Nominal Utterances and Populist Rhetoric in an Italian Twitter Corpus of Hate Speech against Immigrants
Gloria Comandini and Viviana Patti 163

"Condescending, Rude, Assholes": Framing gender and hostility on Stack Overflow
Sian Brooke 172

Online aggression from a sociological perspective: An integrative view on determinants and possible countermeasures
Sebastian Weingartner and Lea Stahel 181