ACL 2017

The First Workshop on Abusive Language Online

Proceedings of the Workshop

August 4, 2017
Vancouver, Canada

Sponsors

Primary Sponsor

Malcolm S. Forbes Center for Culture and Media Studies

Platinum Sponsors

Gold Sponsors

Google amazon

Silver Sponsors

The New York Times Bloomberg

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-945626-66-1

Introduction

We are very pleased to welcome you to the first Workshop on Abusive Language Online (ALW), held at ACL 2017 in Vancouver, Canada. The last few years have seen a surge in abusive behavior online, with governments, social media platforms, and individuals struggling to cope with the consequences and to produce effective methods to combat it. In many cases, online forums, comment sections, and social media interactions have become sites for bullying, scapegoating, and hate speech. These forms of online aggression not only poison the social climate of the online communities that experience it, but can also provoke physical violence and harm.

Addressing abusive language necessitates a multidisciplinary approach that requires knowledge from several fields, including, but not limited to: media studies, natural language processing (NLP), psychology, sociology, law, gender studies, communications, and critical race theory. NLP, as a field that directly works with computationally analyzing language, is in a unique position to develop automated methods to analyse, detect, and filter abusive language. By working across disciplinary divides, researchers in all these fields can produce a comprehensive approach to abusive language that blends together computational, social and legal methods.

We are therefore very happy to bring researchers of various disciplines together in this one-day workshop to discuss approaches to abusive language. The workshop consists of two invited speaker talks, two panels, and oral and poster presentations.

- Carol Todd
  Carol Todd founded the Amanda Todd Legacy in memory of her daughter Amanda after her death by suicide on October 10, 2012. Amanda's Legacy was created to bring increased awareness and conversations within families and communities about online exploitation, cyberabuse and internet safety. The goal has been to encourage a shift in thinking about bullying type behaviours (both on and offline) to those of REFLECTION and RESPECT as well as to destigmatizing the perceptions related mental health as it can relate to how we treat others.

- Brianna Wu
  Brianna Wu is a 2018 Democratic candidate for U.S. Congress in Massachusetts-District 8. Brianna is also head of development at GSX, a Boston independent videogame studio. Brianna came to national prominence when she and other women working in the tech industry were personally targeted by alt-right hate groups, including one spearheaded by Steve Bannon, now Chief Strategist to Donald Trump. Despite threats on her life and her family, Brianna has never wavered as a voice for the marginalized, including women, people of color and LGBT individuals.

We will be hosting the following researchers as our panelists:

- Lucas Dixon
  Lucas Dixon is Chief Scientist at Jigsaw, an incubator within Alphabet that builds technology to tackle

some of the toughest global security challenges facing the world today. His work focuses on security, machine intelligence and data visualization..

- Pascale Fung
  Pascale Fung is a Professor at the Department of Electronic & Computer Engineering at The Hong Kong University of Science & Technology. She is the founding chair of the Women Faculty Association at HKUST and her research interests lies in building intelligent systems that can understand and empathise with humans.

- Sora Han
  Sora Han is an Associate Professor of Criminology, Law and Society at the School of Law at UC Irvine. She recently published her first book, Letters of the Law (Stanford University Press 2015), which recasts and extends the insights of critical race theory to produce new readings of American law's landmark decisions on race and civil rights.

- Elizabeth Losh
  Elizabeth Losh is an Associate Professor of English and American Studies at William and Mary with a specialization in New Media Ecologies. In addition to recent work on selfies and hashtag activism, she has also written a number of frequently cited essays about communities that produce, consume, and circulate online video, videogames, digital photographs, text postings, and programming code.

- Margaret Mitchell
  Margaret Mitchell is the Senior Research Scientist in Google's Research & Machine Intelligence group, working on advancing artificial intelligence towards positive goals. Her work combines computer vision, natural language processing, social media, many statistical methods, and insights from cognitive science.

- Vinodkumar Prabhakaran
  Vinodkumar Prabhakaran is a postdoctoral fellow in the computer science department at Stanford University. His research falls in the interdisciplinary field of computational sociolinguistics, in which he builds and uses computational tools to analyze linguistic patterns that reveal the underlying social contexts in which language is used.

- Jacqueline Wernimont
  Jacqueline is a founding co-Director of the HS Collab and an assistant professor of English at Arizona State University, where she specializes in literary history, feminist digital media, histories of quantification, and technologies of commemoration. Her current book project, tentatively titled Numbered Lives, traces a 500+-year history of technologies that attempt to quantify human life.

In addition, the workshop includes research papers from the community. We received 21 submissions, and accepted 14 (67% acceptance rate): 4 as oral presentations and 10 as poster presentations. For each paper, we assigned three reviewers from within NLP and at least one reviewer from outside of NLP to provide a different perspective on the research. The papers at the workshop cover a wide range of topics: for example, abusive language detection in different languages, analysis of abusive language across different domains, development of corpora and annotation guidelines for this field of NLP, to name a few.

We would like to thank all authors of the submitted papers, reviewers, presenters, invited speakers, and panelists. In addition, we thank our generous sponsors which helped us fund the travel costs for

speakers and panelists: Brown University as our principal sponsor, StrainTek as a platinum sponsor, Google and Amazon as gold sponsors, and the New York Times and Bloomberg as silver sponsors.

It is our hope that this workshop can function as a starting point for more interdisciplinary work, approaches, and cooperation in analyzing and detecting abusive language online.

We wish you all a productive and inspiring workshop!

Zeerak, Wendy, Dirk & Joel

Organizers:

Zeerak Waseem, University of Sheffield
Wendy Hui Kyong Chun, Brown University
Dirk Hovy, University of Copenhagen
Joel Tetreault, Grammerly

Program Committee:

Swati Agarwal, IIIT Delhi, India
Fiona Barnett, Duke University, USA
Darina Benikova, University of Duisburg-Essen, LTL, Germany
Simone Browne, UT Austin, USA
Anneke Buffone, University of Pennsylvania, USA
Pete Burnap, Cardiff University, United Kingdom
Christina Capodilupo, Teachers College, Columbia University, USA
Guillermo Carbonell, University Duisburg-Essen, Germany
Pedro Cardoso, Synthesio, France
Gabriella Coleman, McGill, Canada
Bart Desmet, LT3, Ghent University, Belgium
Lucas Dixon, Jigsaw, USA
Nemanja Djuric, Uber ATC, USA
Jacob Eisenstein, Georgia Institute of Technology, USA
Hugo Jair Escalante, INAOE, Mexico
Lucie Flekova, UKP Lab, TU Darmstadt, Germany
Camille François, Jigsaw, USA
Matthew Fuller, Goldsmith, United Kingdom
Tanton Gibbs, Facebook, USA
Lee Gillam, University of Surrey, United Kingdom
Jen Golbeck, University of Maryland, USA
Erica Greene, New York Times, USA
Kevin Hamilton, University of Illinois, USA
Sora Han, University of California, Irvine, USA
Christopher Homan, Rochester Institute of Technology, USA
Veronique Hoste, Ghent University, Belgium
Ruihong Huang, Texas A&M, USA
Els Lefever, LT3, Ghent University, Belgium
Shuhua Liu, Arcada University of Applied Sciences, Finland
Elizabeth Losh, College of William and Mary, USA
Shervin Malmasi, Harvard Medical School, USA
Fumito Masui, Kitami Institute of Technology, Japan
Yashar Mehdad, Airbnb, USA
Rada Mihalcea, University of Michigan, USA
Mainack Mondal, Max Planck Institute for Software Systems, Germany
Manuel Montes-y-Gómez, INAOE, Mexico
Kevin Munger, NYU, USA
Srmuthi Mukund, A9.com Inc, USA
Preslav Nakov, Qatar Computing Research Institute, HBKU, Qatar

Courtney Napoles, Johns Hopkins University, USA
Chikashi Nobata, Apple, USA
Guy De Pauw, CLiPS - University of Antwerp, Belgium
Whitney Phillips, Mercer University, USA
Karolien Poels, University of Antwerp, Belgium
Daniel Preotiuc-Pietro, University of Pennsylvania, USA
Michal Ptaszynski, Kitami Institute of Technology, Japan
Awais Rashid, Lancaster University, United Kingdom
Björn Ross, University Duisburg-Essen, Germany
Paolo Rosso, Universitat Politecnica de Valencia, Spain
Masoud Rouhizadeh, Stony Brook University & University of Pennsylvania, USA
Christina Sauper, Facebook, USA
Molly Sauter, McGill University, Canada
Nishant Shah, Leuphana, ArtEZ University of the Arts, CIS (Bangalore), India
Thamar Solorio, University of Houston, USA
Jeffrey Sorensen, Jigsaw, USA
Dennis Tenen, Columbia University, USA
Jennifer Terry, University of California, Irvine, USA
Achint Thomas, Embibe Indiavidual Inc, India
Nanna Bonde Thylstrup, University of Copenhagen, Denmark
Lyle Ungar, University of Pennsylvania, USA
Anna Vartapetiance, University of Surrey, United Kingdom
Kristin Veel, University of Copenhagen, Denmark
Erik Velldal, University of Oslo, Norway
Ingmar Weber, Qatar Computing Research Institute, Qatar
Jacque Wernimont, Arizona State University, USA
Michael Wojatzki, University of Duisburg-Essen, Germany
Lilia Øvrelid, University of Oslo, Norway

Invited Speakers:

Brianna Wu, Giant Spacekat, USA
Carol Todd, Amanda Todd Legacy Society, Canada

Panelists:

Lucas Dixon, Jigsaw, USA
Pascale Fung, Hong Kong University of Science and Technology, Hong Kong
Sora Han, University of California, Irvine, USA
Elizabeth Losh, William and Mary, USA
Margaret Mitchell, Google, USA
Vinodkumar Prabhakaran, Stanford University, USA
Jacqueline Wernimont, Arizona State University, USA

Table of Contents

Conference Program

Friday, August 4, 2017

08:45–09:05 ***Opening Remarks***

09:05–09:50 ***Invited Talk A: Carol Todd***

09:50–10:35 ***Panel A: Sora Han, Liz Losh, Lucas Dixon***

10:35–11:00 ***Break***

11:00–12:30 ***Paper Presentations***

11:00–11:20 *Dimensions of Abusive Language on Twitter*
Isobelle Clarke and Dr. Jack Grieve

11:20–11:40 *Constructive Language in News Comments*
Varada Kolhatkar and Maite Taboada

11:40–12:00 *Rephrasing Profanity in Chinese Text*
Hui-Po Su, Zhen-Jie Huang, Hao-Tsung Chang and Chuan-Jie Lin

12:00–12:20 *Deep Learning for User Comment Moderation*
John Pavlopoulos, Prodromos Malakasiotis and Ion Androutsopoulos

12:20–14:00 ***Lunch***

14:00–15:30 ***Poster Session***

Class-based Prediction Errors to Detect Hate Speech with Out-of-vocabulary Words
Joan Serrà, Ilias Leontiadis, Dimitris Spathis, Gianluca Stringhini, Jeremy Blackburn and Athena Vakali

One-step and Two-step Classification for Abusive Language Detection on Twitter
Ji Ho Park and Pascale Fung

Friday, August 4, 2017 (continued)

Legal Framework, Dataset and Annotation Schema for Socially Unacceptable Online Discourse Practices in Slovene
Darja Fišer, Tomaž Erjavec and Nikola Ljubešić

Abusive Language Detection on Arabic Social Media
Hamdy Mubarak, Kareem Darwish and Walid Magdy

Vectors for Counterspeech on Twitter
Lucas Wright, Derek Ruths, Kelly P Dillon, Haji Mohammad Saleem and Susan Benesch

Detecting Nastiness in Social Media
Niloofar Safi Samghabadi, Suraj Maharjan, Alan Sprague, Raquel Diaz-Sprague and Thamar Solorio

Technology Solutions to Combat Online Harassment
George Kennedy, Andrew McCollough, Edward Dixon, Alexei Bastidas, John Ryan, Chris Loo and Saurav Sahay

Understanding Abuse: A Typology of Abusive Language Detection Subtasks
Zeerak Waseem, Thomas Davidson, Dana Warmsley and Ingmar Weber

Using Convolutional Neural Networks to Classify Hate-Speech
Björn Gambäck and Utpal Kumar Sikdar

Illegal is not a Noun: Linguistic Form for Detection of Pejorative Nominalizations
Alexis Palmer, Melissa Robinson and Kristy K. Phillips

15:30–16:00 ***Break***

16:00–16:45 ***Invited Talk B: Brianna Wu***

16:45–17:30 ***Panel B: Pascale Fung, Vinodkumar Prabhakaran, Jacqueline Wernimont, Margeret Mitchell***

17:30–17:40 ***Wrapup***