ALW2: 2nd Workshop on Abusive Language Online
EMNLP 2018 (Brussels, Belgium), October 31st or November 1st, 2018
Submission deadline: July 27th, 2018
Website: https://sites.google.com/view/alw2018
Submission link: https://www.softconf.com/emnlp2018/ALW2/

***** Deadline Extended: New Submission deadline July 27th *****
***** Accepted papers will be considered for a Special issue in the journal First Monday (http://firstmonday.org) planned for Fall 2019 *****

Overview

Interaction amongst users on social networking platforms can enable constructive and insightful conversations and civic participation; however, on many sites that encourage user interaction, verbal abuse has become commonplace, leading to negative outcomes such as cyberbullying, hate speech, and scapegoating. In online contexts, aggressive behavior may be more frequent than in face-to-face interaction, which can poison the social climates within online communities. The last few years have seen a surge in such abusive online behavior, leaving governments, social media platforms, and individuals struggling to deal with the consequences.

For instance, in 2015, Twitter's CEO publicly admitted that online abuse on their platform was resulting in users leaving the platform, and in some cases even having to leave their homes. More recently, Facebook, Twitter, YouTube and Microsoft pledged to remove hate speech from their platforms within 24 hours in accordance with the EU commission code of conduct and face fines of up to €50M in Germany if they systematically fail to remove abusive content within 24 hours. While governance demands the ability to respond quickly and at scale, we do not yet have effective human or technical processes that can address this need. Abusive language can often be extremely subtle and highly context dependent. Thus we are challenged to develop scalable computational methods that can reliably and efficiently detect and mitigate the use of abusive language online within variable and evolving contexts.

As a field that works directly with computational analysis of language, NLP (Natural Language Processing) is in a unique position to address this problem. Recently there have been a greater number of papers dealing with abusive language in the computational linguistics community. Abusive language is not a stable or simple target: misclassification of regular conversation as abusive can severely impact users' freedom of expression and reputation, while misclassification of abusive conversations as unproblematic on the other hand maintains the status quo of online communities as unsafe environments. Clearly, there is still a great deal of work to be done in this area. More practically, as research into detecting abusive language is still in its infancy, the research community has yet to agree upon a suitable typology of abusive content as well as upon standards and metrics for proper evaluation, where research in media studies, rhetorical analysis, and cultural analysis can offer many insights.

In this second edition of this workshop, we continue to emphasize the computational detection of abusive language as informed by interdisciplinary scholarship and community experience. We invite paper submissions describing unpublished work from relevant fields including, but not limited to: natural language processing, law, psychology, network analysis, gender and women's studies, and critical race theory.

Paper Topics

We invite long and short papers on any of the following **general topics**:

related to developing computational models and systems:

- NLP models and methods for detecting abusive language online, including, but not limited to hate speech, cyberbullying etc.
- Application of NLP tools to analyze social media content and other large data sets
- NLP models for cross-lingual abusive language detection
- Computational models for multi-modal abuse detection
- Development of corpora and annotation guidelines
- Critical algorithm studies with a focus on abusive language moderation technology
- Human-Computer Interaction for abusive language detection systems
- Best practices for using NLP techniques in watchdog settings

or related to legal, social, and policy considerations of abusive language online:

- The social and personal consequences of being the target of abusive language and targeting others with abusive language
- Assessment of current non-NLP methods of addressing abusive language
- Legal ramifications of measures taken against abusive language use
- Social implications of monitoring and moderating unacceptable content
- Considerations of implemented and proposed policies for dealing with abusive language online and the technological means of dealing with it.

In addition, in this one-day workshop, we will have

1. a multidisciplinary panel discussion and
2. a forum for plenary discussion on the issues that researchers and practitioners face in efforts to work with abusive language detection and
3. selected submissions from the workshop will be published in a special issue in the journal First Monday.

We seek to have a greater focus on policy aspects of online abuse through invited speakers and panels.

Unshared task

In order to encourage focused contributions, we encourage researchers to consider using one or more of the following datasets in their experiments:

- StackOverflow Offensive Comments [Access from workshop webpage]
- Yahoo News Dataset of User Comments [Nobata et al., WWW 2016]
- Twitter Data Set [Waseem and Hovy, NAACL 2016]
- German Twitter Data Set [Ross et al. NLP4CMC 2016]
- Greek News Data Set [Pavlopoulos et al., EMNLP 2017]
- Wikimedia Toxicity Data Set [Wulczyn et al., WWW 2017]
- SFU Opinion and Comment Corpus [Kolhatkar et al., In Review]
- Conversations Gone Awry [Zhang et al., ACL 2018]

Submission Information

We will be using the EMNLP 2018 Submission Guidelines. Authors are invited to submit a **full paper** of up to **8 pages** of content with up to 2 additional pages for references. We also invite **short papers** of up to **4 pages** of content, including 2 additional pages for references. Accepted papers will be given an additional page of content to address reviewer comments. We also invite papers which describe systems. If you would like to present a **demo** in addition to presenting the paper, please make sure to select either "full paper + demo" or "short paper + demo" under "Submission Category" in the START submission page.

Previously published papers cannot be accepted. The submissions will be reviewed by the program committee. As reviewing will be blind, please ensure that papers are anonymous. Self-references that reveal the author's identity, e.g., "We previously showed (Smith, 1991) ...", should be avoided. Instead, use citations such as "Smith previously showed (Smith, 1991) ...".

We have also included **conflict of interest** in the submission form. You should mark all potential reviewers who have been authors on the paper, are from the same research group or institution, or who have seen versions of this paper or discussed it with you.

We will be using the START conference system to manage submissions.

Important Dates

Submission due: July 27, 2018
Author Notification: August 18, 2018
Camera Ready: August 31, 2018
Workshop Date: Oct 31st or Nov 1st, 2018
Submission link: https://www.softconf.com/emnlp2018/ALW2/

Organizing Committee

- Darja Fišer, University of Ljubljana & the Jožef Stefan Institute
- Ruihong Huang, Texas A&M University
- Vinodkumar Prabhakaran, Stanford University
- Rob Voigt, Stanford University
- Zeerak Waseem, University of Sheffield
- Jacqueline Wernimont, Arizona State University

Program Committee/Reviewers

The following researchers have agreed to serve on the program committee as reviewers.

- Ion Androutsopoulos, Athens University of Economics and Business, Greece

- Veronika Bajt, Peace Institute, Slovenia
- Susan Benesch, Dangerous Speech Project, United States
- Darina Benikova, University of Duisburg-Essen, Germany
- Joachim Bingel, University of Copenhagen, Denmark
- Anne Brigitta Nilsen, Oslo Metropolitan University, Norway
- Wendy Chun, Brown University, United States
- Kelly Dennis, University of Connecticut, United States
- Lucas Dixon, Jigsaw/Google, United States
- Nemanja Djuric, Uber ATG, United States
- Micha Elsner, The Ohio State University, United States
- Hugo Jair Escalante, INAOE, Mexico
- Björn Gambäck, Norwegian University of Science and Technology, Norway
- Lee Gillam, University of Surrey, United Kingdom
- Tassie Gniady, Indiana University, United States
- Vojko Gorjanc, University of Ljubljana, Slovenia
- Erica Greene, Jigsaw/Google, United States
- Julia Hockenmaier, University of Illinois Urbana-Champaign, United States
- Veronique Hoste, Ghent University, Belgium
- Dirk Hovy, Bocconi University, Italy
- Dan Jurafsky, Stanford University, United States
- George Kennedy, Intel, United States
- Neža Kogovšek Šalomon, Peace Institute, Slovenia
- Els Lefever, LT3, Ghent University, Belgium
- Chuan-Jie Lin, National Taiwan Ocean Universty, Taiwan
- Elizabeth Losh, William and Mary, United States
- Prodromos Malakasiotis, StrainTek, Athens University of Economics and Business Informatics Department, Greece
- Shervin Malmasi, Harvard Medical School, United States
- Diana Maynard, University of Sheffield, United Kingdom
- Kathy McKeown, Columbia University, United States
- Mainack Mondal, University of Chicago, United States
- Hamdy Mubarak, Qatar Computing Research Institute, Qatar Foundation, Qatar
- smruthi mukund, Amazon, United States
- Kevin Munger, NYU, United States
- Preslav Nakov, Qatar Computing Research Institute, HBKU, Qatar
- Chikashi Nobata, Apple Inc., United States
- Gustavo Paetzold, Federal University of Technology, Brazil
- John Pavlopoulos, Athens University of Economics and Business, Greece

- Daniel Preoțiuc-Pietro, University of Pennsylvania, United States
- Michal Ptaszynski, Kitami Institute of Technology, Japan
- Preethi Raghavan, IBM Research TJ Watson, United States
- Georg Rehm, DFKI, Germany
- Björn Ross, University of Duisburg-Essen, Germany
- Masoud Rouhizadeh, Johns Hopkins University, United States
- Niloofar Safi Samghabadi, University of Houston, United States
- Christina Sauper, Facebook, United States
- Alexandra Schofield, Cornell University, United States
- Caroline Sinders, Wikimedia Foundation, United States
- Maite Taboada, Simon Fraser University, Canada
- Dennis Tenen, Columbia University, United States
- Joris Van Hoboken, Vrije Universiteit Brussel, Belgium
- Ingmar Weber, Qatar Computing Research Institute, Qatar
- Amanda Williams, University of Bristol, United Kingdom
- Michael Wojatzki, Language Technology Lab, University of Duisburg-Essen, Germany
- Lilja Øvrelid, Dept of Informatics, University of Oslo, Norway

Related Events

- Workshop: The turn to artificial intelligence in governing communication online
- First Workshop on Trolling, Aggression and Cyberbullying
- The 1st Workshop on Abusive Language Online: the first edition of the workshop.
- CHI Workshop on Online Harassment: a workshop focused on developing datasets for researching online harassment
- Text Analytics for Cyber Security and Online Safety, LREC 2016
- Discourses of Aggression and Violence in Greek Digital Communication, ICGL13
- Conceptualizing, Creating, & Controlling Constructive and Controversial Comments: A CSCW Research-athon