

LREC 2020 Workshop
Language Resources and Evaluation Conference
11–16 May 2020

**Second Workshop on
Trolling, Aggression and Cyberbullying
(TRAC-2, 2020)**

PROCEEDINGS

Editors:

Ritesh Kumar, Atul Kr. Ojha, Bornini Lahiri, Marcos Zampieri,
Shervin Malmasi, Vanessa Murdock and Daniel Kadar

Proceedings of the LREC 2020 Second Workshop on Trolling, Aggression and Cyberbullying (TRAC-2, 2020)

Edited by:

Ritesh Kumar, Atul Kr. Ojha, Bornini Lahiri, Marcos Zampieri,
Shervin Malmasi, Vanessa Murdock and Daniel Kadar

ISBN: 979-10-95546-56-6

EAN: 9791095546566

For more information:

European Language Resources Association (ELRA)

9 rue des Cordelières

75013, Paris

France

<http://www.elra.info>

Email: lrec@elda.org

© European Language Resources Association (ELRA)

These workshop proceedings are licensed under a Creative Commons
Attribution-NonCommercial 4.0 International License

Introduction

In the last few years, we have witnessed a gradual shift from largely static, read-only web to quickly expanding user-generated web. There has been an exponential growth in the availability and use of online platforms where users can post their own content. A major part of these platforms include social media websites and apps, blogs, Q&A forums and several similar platforms. All of these are almost exclusively user-generated websites. In all of these platforms and forums, humongous amount of data is created and circulated every minute. It has been estimated that there has been an increase of approximately 25% in the number of tweets per minutes and 22% increase in the number of Facebook posts per minute in the last 3 years. It is posited that approximately 500 million tweets are sent per day, 4.3 billion Facebook messages are posted and more than 200 million emails are sent each day, and approximately 2 million new blog posts are created daily over the web ¹. There is no such thing as a ‘consolidated figure’ of the number of comments and opinion generated on websites worldwide, but it can be safely assumed that such a figure would be staggering.

As the number of people and this interaction over the web has increased, incidents of aggression and related activities like trolling, cyberbullying, flaming, hate speech, etc. have also increased manifold across the globe. The reach and extent of Internet has given such incidents unprecedented power and influence to affect the lives of billions of people. It has been reported that such incidents of online abuse have not only created mental and psychological health issues for users, but they have impacted our lives in many other way, spanning from deactivating accounts to instances of self-harm and suicide. NLP and related methods have shown great promise in dealing with such abusive behaviour through early detection of inflammatory content.

This workshop focusses on the phenomena of online aggression, trolling, cyberbullying and other related phenomena, in both text (especially social media) and speech. The organisers aim to create a platform for academic discussions on this phenomena, based on previous joint work that they have done as part of a project funded by the British Council. We are particularly interested in promoting conversations dedicated to the automatic detection of aggression in both speech and text, that is, we hope that our workshop will not only be purely academic by nature but it will also generate real-life solutions to tackle the phenomena studied. As such the workshop also includes a shared task on ‘Aggression Identification’. The task consisted of two sub-tasks - aggression identification (sub-task A) and gendered identification (sub-task B) - in three languages - Bangla, Hindi and English. For this task, the participants were provided with a dataset of approximately 5,000 instances from YouTube comments in each language. Additional data for testing was released at a later date.

Both the workshop and the shared task received a very encouraging response from the community. There were more than 70 registrations for the shared task. Out of these, 19 teams submitted their systems. The proceedings include 13 system description papers that were finally submitted by the authors. In addition to this, the workshop also includes 16 regular papers presented in the workshop.

We would like to thank all the authors for their submission and members of the Program Committee for their invaluable efforts in reviewing and providing feedback to all the papers. We would also like to thank all the members of the Organising Committee who have helped immensely in various aspects of the organisation of the workshop and the shared task.

¹Source: <https://www.gwava.com/blog/internet-data-created-daily/>

Workshops Chairs

Ritesh Kumar, Dr. Bhimrao Ambedkar University, India
Atul Kr. Ojha, Charles University, Prague & Panlingua Language Processing LLP, India
Bornini Lahiri, Indian Institute of Technology-Kharagpur, India

Organising Committee

Ritesh Kumar, Dr. Bhimrao Ambedkar University, India
Atul Kr. Ojha, Charles University, Prague & Panlingua Language Processing LLP, India
Vanessa Murdock, Amazon Inc., USA
Marcos Zampieri, Rochester Institute of Technology, USA
Shervin Malmasi, Amazon Inc., USA
Bornini Lahiri, Indian Institute of Technology-Kharagpur, India
Daniel Kadar, Research Institute for Linguistics, Hungarian Academy of Sciences, Hungary

Shared Task Organising Committee

Ritesh Kumar, Dr. Bhimrao Ambedkar University, India
Atul Kr. Ojha, Charles University, Prague & Panlingua Language Processing LLP, India
Marcos Zampieri, Rochester Institute of Technology, USA
Shervin Malmasi, Harvard Medical School, USA

Editors

Ritesh Kumar, Dr. Bhimrao Ambedkar University, India
Atul Kr. Ojha, Charles University, Prague & Panlingua Language Processing LLP, India
Bornini Lahiri, Indian Institute of Technology-Kharagpur, India
Marcos Zampieri, Rochester Institute of Technology, USA
Shervin Malmasi, Amazon Inc., USA
Vanessa Murdock, Amazon Inc., USA
Daniel Kadar, Research Institute for Linguistics, Hungarian Academy of Sciences, Hungary

Programme Committee

A. Seza Dođruöz, Tilburg University, Netherlands
Adrián Pastor López Monroy, University of Houston, USA
Amitava Das, WIPRO Limited, India
Asif Ekbal, IIT-Patna, India
Atul Kr. Ojha, Charles University, Prague & Panlingua Language Processing LLP, India
Bornini Lahiri, Indian Institute of Technology-Kharagpur, India
Bruno Emanuel Martins, University of Lisbon, Portugal
Cheng-Te Li, National Cheng Kung University, Taiwan
Chuan-Jie Lin, National Taiwan Ocean University, Taiwan
Claudia Peersman, Lancaster University, UK
Cynthia van Hee, LT3, Ghent University, Belgium
Danilo Croce, University of Roma, Italy
Dennis Tenen, Columbia University, USA
Elizabeth Losh, William and Mary College, USA
Els Lefever, LT3, Ghent University, Belgium
Erik Velldal, University of Oslo, Norway
Eshwar Chandresekharan, Georgia Tech, USA
Fumito Masui, Kitami Institute of Technology, Japan
Girish Nath Jha, Jawaharlal Nehru University, India
Haris Papageorgiou, ATHENA Research and Innovation Center, Greece
Hugo Jair Escalante, INAOE, Mexico
Ingmar Weber, Qatar Computing Research Institute, Qatar
Jen Golbeck, University of Maryland, USA
Jacqueline Wernimont, Arizona State University, USA
Kalika Bali, MSRI Bangalore, India
Lee Gillam, University of Surrey, UK
Liang-Chih Yu, Yuan Ze University, Taiwan
Libby Hemphill, University of Michigan, USA
Lun-Wei Ku, Academia Sinica, Taiwan
Mainack Mondal, University of Chicago, USA
Manuel Montes-y-Gómez, INAOE, Mexico
Marco Guerini, Fondazione Bruno Kessler, Trento, Italy
Marcos Zampieri, Rochester Institute of Technology, USA
Matthew Fuller, University of London, UK
Michael Wiegand, Saarland University, Germany
Michael Paul, University of Colorado Boulder, USA
Min-Yuh Day, Tamkong University, Taiwan
Ming-Feng Tsai, National Chengchi University, Taiwan
Monojit Choudhury, MSRI Bangalore, India
Michal Ptaszynski, Kitami Institute of Technology, Japan
Nemanja Djuric, Uber ATC, USA
Pawan Goyal, IIT-Kharagpur, India
Pete Burnap, Cardiff University, UK
Preslav Nakov, Qatar Computing Research Institute, Qatar

Ritesh Kumar, Dr. B. R. Ambedakar University, India
Roman Klinger, University of Stuttgart, Germany
Ruifeng Xu, Harbin Institute of Technology, China
Saptarshi Ghosh, IIT-Kharagpur, India
Sara E. Garza, Universidad Autónoma de Nuevo León (UANL), Mexico
Shervin Malmasi, Amazon Inc., USA
Thamar Solorio, University of Houston, USA
Thiago Galery, DBpedia Association and Idio, London, UK
Vanessa Murdock, Amazon Inc., USA
Veronique Hoste, LT3, Ghent University, Belgium
Vladan Radosavljevic, Temple University, USA
William Wang, University of California-Santa Barbara, USA
Xavier Tannier, Université Paris-Sud, LIMSI, CNRS, France
Ye Tian, Amazon Cambridge Development Centre (Alexa), UK
Yelena Mejova, Qatar Computing Research Institute, Qatar
Zeeraq Waseem, University of Sheffield, UK
Zhunchen Luo, China Defense Science and Technology Information Center, China

Table of Contents

<i>Evaluating Aggression Identification in Social Media</i> Ritesh Kumar, Atul Kr. Ojha, Shervin Malmasi and Marcos Zampieri	1
<i>TOCP: A Dataset for Chinese Profanity Processing</i> Hsu Yang and Chuan-Jie Lin	6
<i>A Multi-Dimensional View of Aggression when voicing Opinion</i> Arjit Srivastava, Avijit Vajpayee, Syed Sarfaraz Akhtar, Naman Jain, Vinay Singh and Manish Shrivastava	13
<i>Towards Non-Toxic Landscapes: Automatic Toxic Comment Detection Using DNN</i> Ashwin Geet D'Sa, Irina Illina and Dominique Fohr	21
<i>Aggression Identification in Social Media: a Transfer Learning Based Approach</i> Faneva Ramiandrisoa and Josiane Mothe	26
<i>Multimodal Meme Dataset (MultiOFF) for Identifying Offensive Content in Image and Text</i> Shardul Suryawanshi, Bharathi Raja Chakravarthi, Mihael Arcan and Paul Buitelaar	32
<i>A Comparative Study of Different State-of-the-Art Hate Speech Detection Methods in Hindi-English Code-Mixed Data</i> Priya Rani, Shardul Suryawanshi, Koustava Goswami, Bharathi Raja Chakravarthi, Theodorus Franssen and John Philip McCrae	42
<i>IRIT at TRAC 2020</i> Faneva Ramiandrisoa and Josiane Mothe	49
<i>Bagging BERT Models for Robust Aggression Identification</i> Julian Risch and Ralf Krestel	55
<i>Scmhl5 at TRAC-2 Shared Task on Aggression Identification: Bert Based Ensemble Learning Approach</i> Han Liu, Pete Burnap, Wafa Alorainy and Matthew Williams	62
<i>The Role of Computational Stylometry in Identifying (Misogynistic) Aggression in English Social Media Texts</i> Antonio Pascucci, Raffaele Manna, Vincenzo Masucci and Johanna Monti	69
<i>Aggression Identification in English, Hindi and Bangla Text using BERT, RoBERTa and SVM</i> Arup Baruah, Kaushik Das, Ferdous Barbhuiya and Kuntal Dey	76
<i>LaSTUS/TALN at TRAC - 2020 Trolling, Aggression and Cyberbullying</i> Lütfiye Seda Mut Altın, Alex Bravo and Horacio Saggion	83
<i>Spyder: Aggression Detection on Multilingual Tweets</i> Anisha Datta, Shukrity Si, Urbi Chakraborty and Sudip Kumar Naskar	87
<i>BERT of all trades, master of some</i> Denis Gordeev and Olga Lykova	93
<i>SAJA at TRAC 2020 Shared Task: Transfer Learning for Aggressive Identification with XGBoost</i> Saja Tawalbeh, Mahmoud Hammad and Mohammad AL-Smadi	99

<i>FlorUniTo@TRAC-2: Retrofitting Word Embeddings on an Abusive Lexicon for Aggressive Language Detection</i>	
Anna Koufakou, Valerio Basile and Viviana Patti	106
<i>AI_ML_NIT_Patna @ TRAC - 2: Deep Learning Approach for Multi-lingual Aggression Identification</i>	
Kirti Kumari and Jyoti Prakash Singh	113
<i>Multilingual Joint Fine-tuning of Transformer models for identifying Trolling, Aggression and Cyberbullying at TRAC 2020</i>	
Sudhanshu Mishra, Shivangi Prasad and Shubhanshu Mishra	120
<i>Aggression and Misogyny Detection using BERT: A Multi-Task Approach</i>	
Niloofer Safi Samghabadi, Parth Patwa, Srinivas PYKL, Prerana Mukherjee, Amitava Das and Thamar Solorio	126
<i>Automatic Detection of Offensive Language in Social Media: Defining Linguistic Criteria to build a Mexican Spanish Dataset</i>	
María José Díaz-Torres, Paulina Alejandra Morán-Méndez, Luis Villasenor-Pineda, Manuel Montesy-Gómez, Juan Aguilera and Luis Meneses-Lerín	132
<i>Offensive Language Detection Explained</i>	
Julian Risch, Robin Ruff and Ralf Krestel	137
<i>Detecting Early Signs of Cyberbullying in Social Media</i>	
Niloofer Safi Samghabadi, Adrián Pastor López Monroy and Thamar Solorio	144
<i>Lexicon-Enhancement of Embedding-based Approaches Towards the Detection of Abusive Language</i>	
Anna Koufakou and Jason Scott	150
<i>Developing a Multilingual Annotated Corpus of Misogyny and Aggression</i>	
Shiladitya Bhattacharya, Siddharth Singh, Ritesh Kumar, Akanksha Bansal, Akash Bhagat, Yogesh Dawer, bornini lahiri and Atul Kr. Ojha	158

Conference Program

Saturday, May 16, 2020

9:00–10:30 **Inaugural Session**

9:00–9:10 *Welcome by Workshop Chairs*

9:10–9:30 *Evaluating Aggression Identification in Social Media*
Ritesh Kumar, Atul Kr. Ojha, Shervin Malmasi and Marcos Zampieri

9:30–10:30 **Paper Session I**

9:30–9:50 *TOCP: A Dataset for Chinese Profanity Processing*
Hsu Yang and Chuan-Jie Lin

9:50–10:10 *A Multi-Dimensional View of Aggression when voicing Opinion*
Arjit Srivastava, Avijit Vajpayee, Syed Sarfaraz Akhtar, Naman Jain, Vinay Singh
and Manish Shrivastava

10:10–10:30 *Towards Non-Toxic Landscapes: Automatic Toxic Comment Detection Using DNN*
Ashwin Geet D'Sa, Irina Illina and Dominique Fohr

10:30–10:45 *Break*

Saturday, May 16, 2020 (continued)

10:45–11:45 Paper Session II

10:45–11:05 *Aggression Identification in Social Media: a Transfer Learning Based Approach*

Faneva Ramiandrisoa and Josiane Mothe

11:05–11:25 *Multimodal Meme Dataset (MultiOFF) for Identifying Offensive Content in Image and Text*

Shardul Suryawanshi, Bharathi Raja Chakravarthi, Mihael Arcan and Paul Buitelaar

11:25–11:45 *A Comparative Study of Different State-of-the-Art Hate Speech Detection Methods in Hindi-English Code-Mixed Data*

Priya Rani, Shardul Suryawanshi, Koustava Goswami, Bharathi Raja Chakravarthi, Theodorus Fransen and John Philip McCrae

11:45–14:00 Poster Session

IRIT at TRAC 2020

Faneva Ramiandrisoa and Josiane Mothe

Bagging BERT Models for Robust Aggression Identification

Julian Risch and Ralf Krestel

Scmh15 at TRAC-2 Shared Task on Aggression Identification: Bert Based Ensemble Learning Approach

Han Liu, Pete Burnap, Wafa Alorainy and Matthew Williams

The Role of Computational Stylometry in Identifying (Misogynistic) Aggression in English Social Media Texts

Antonio Pascucci, Raffaele Manna, Vincenzo Masucci and Johanna Monti

Aggression Identification in English, Hindi and Bangla Text using BERT, RoBERTa and SVM

Arup Baruah, Kaushik Das, Ferdous Barbhuiya and Kuntal Dey

LaSTUS/TALN at TRAC - 2020 Trolling, Aggression and Cyberbullying

Lütfiye Seda Mut Altın, Alex Bravo and Horacio Saggion

Spyder: Aggression Detection on Multilingual Tweets

Anisha Datta, Shukrity Si, Urbi Chakraborty and Sudip Kumar Naskar

Saturday, May 16, 2020 (continued)

BERT of all trades, master of some

Denis Gordeev and Olga Lykova

SAJA at TRAC 2020 Shared Task: Transfer Learning for Aggressive Identification with XGBoost

Saja Tawalbeh, Mahmoud Hammad and Mohammad AL-Smadi

FlorUniTo@TRAC-2: Retrofitting Word Embeddings on an Abusive Lexicon for Aggressive Language Detection

Anna Koufakou, Valerio Basile and Viviana Patti

AI_ML_NIT_Patna @ TRAC - 2: Deep Learning Approach for Multi-lingual Aggression Identification

Kirti Kumari and Jyoti Prakash Singh

Multilingual Joint Fine-tuning of Transformer models for identifying Trolling, Aggression and Cyberbullying at TRAC 2020

Sudhanshu Mishra, Shivangi Prasad and Shubhanshu Mishra

Aggression and Misogyny Detection using BERT: A Multi-Task Approach

Nilofar Safi Samghabadi, Parth Patwa, Srinivas PYKL, Prerana Mukherjee, Amitava Das and Thamar Solorio

13:00–14:00 *Lunch Break*

14:00–14:45 *Keynote Talk*
Leon Derczynski

Saturday, May 16, 2020 (continued)

14:45–15:45 Paper Session III

14:45–15:05 *Automatic Detection of Offensive Language in Social Media: Defining Linguistic Criteria to build a Mexican Spanish Dataset*

María José Díaz-Torres, Paulina Alejandra Morán-Méndez, Luis Villasenor-Pineda, Manuel Montes-y-Gómez, Juan Aguilera and Luis Meneses-Lerín

15:05–15:25 *Offensive Language Detection Explained*

Julian Risch, Robin Ruff and Ralf Krestel

15:25–15:45 *Detecting Early Signs of Cyberbullying in Social Media*

Nilofar Safi Samghabadi, Adrián Pastor López Monroy and Thamar Solorio

15:45–16:00 Break

16:00–17:00 Paper Session IV

16:00–16:20 *Online Harassment Detection and Categorization using Deep Learning Techniques*
Andrei Paraschiv and CERCEL DUMITRU CLEMENTIN

16:20–16:40 *Lexicon-Enhancement of Embedding-based Approaches Towards the Detection of Abusive Language*

Anna Koufakou and Jason Scott

16:40–17:00 *Developing a Multilingual Annotated Corpus of Misogyny and Aggression*

Shiladitya Bhattacharya, Siddharth Singh, Ritesh Kumar, Akanksha Bansal, Akash Bhagat, Yogesh Dawer, bornini lahiri and Atul Kr. Ojha

Saturday, May 16, 2020 (continued)

17:00–17:10 **Closing**

17:00–17:10 *Vote of Thanks*