



LREC 2022 Workshop
Language Resources and Evaluation Conference
20-25 June 2022

**International Workshop on Resources and Techniques for
User Information in Abusive Language Analysis
(ResT-UP 2)**

PROCEEDINGS

Editors:

Johanna Monti, Valerio Basile, Maria Pia Di Buono, Raffaele
Manna, Antonio Pascucci, Sara Tonelli

Proceedings of the LREC 2022 workshop on Resources and Techniques for User Information in Abusive Language Analysis (ResT-UP 2022)

Edited by: Johanna Monti, Valerio Basile, Maria Pia Di Buono, Raffaele Manna, Antonio Pascucci, Sara Tonelli

ISBN: 979-10-95546-99-3

EAN: 9791095546993

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

Preface by the Workshop Organizers

Welcome to the LREC2022 Workshop on Resources and Techniques for User Information in Abusive Language Analysis (ResT-UP 2).

This volume documents the Proceedings of the Second International Workshop on Resources and Techniques for User Information in Abusive Language Analysis (ResT-UP 2), held on 24 June 2022 in Marseille (France) in conjunction with the LREC 2022 conference (International Conference on Language Resources and Evaluation).

The workshop aims at bringing together researchers and scholars working on author profiling and automatic detection of abusive language on the Web, e.g., cyberbullying or hate speech, with a twofold objective: improving existing LRs, e.g., datasets, corpora, lexicons, and sharing ideas on stylometry techniques and features useful to characterize abusive language online in a fair and transparent way. ResT-UP 2 targets Profiling scholars and research groups, experts in Statistical and Stylistic Analysis of texts as well as computational linguists who investigate author profiling and personality both in short texts (social media posts, blog texts and email) and in long texts (such as pamphlets, (fake) news and political documents). ResT-UP 2 represented an opportunity to share profiling experiments with the scientific community and to show automatic detection techniques of abusive language on the Web.

The workshop has been held as a hybrid event: the workshop was attended by about fifty people (onsite and remotely) between workshop organizers, panelists, keynote speakers and representatives of academic and industrial organizations.

The programme included four oral presentations and two keynote speakers: Viviana Patti and Walter Daelemans.

We would like to thank the invited speakers, all authors who contributed papers to this workshop edition, the Programme Committee members who provided valuable feedback during the review process and the LREC 2022 conference organizing committee.

Johanna Monti – L’Orientale University of Naples – UNIOR NLP Research Group
Valerio Basile – University of Turin – Content-centered Computing group
Maria Pia Di Buono – L’Orientale University of Naples – UNIOR NLP Research Group
Raffaele Manna – L’Orientale University of Naples – UNIOR NLP Research Group
Antonio Pascucci – L’Orientale University of Naples – UNIOR NLP Research Group
Sara Tonelli – Fondazione Bruno Kessler – Digital Humanities research group

- **Organizers:**

Johanna Monti – L’Orientale University of Naples – UNIOR NLP Research Group
Valerio Basile – University of Turin - Content-centered Computing group
Maria Pia Di Buono – L’Orientale University of Naples – UNIOR NLP Research Group
Raffaele Manna – L’Orientale University of Naples – UNIOR NLP Research Group
Antonio Pascucci – L’Orientale University of Naples – UNIOR NLP Research Group
Sara Tonelli – Fondazione Bruno Kessler, Digital Humanities research group

- **Program Committee:**

Cristina Bosco, University of Turin (ITALY)
Maciej Eder, Institute of Polish Language (Polish Academy of Sciences) (POLAND)
Francesca Frontini, Istituto di Linguistica Computazionale "A. Zampolli" - CNR & CLARIN ERIC (ITALY)
Stefano Menini, Fondazione Bruno Kessler (ITALY)
Suzanne Mpouli, Université de Paris (FRANCE)
Michael Oakes, University of Wolverhampton (UNITED KINGDOM)
Alessio Palmero Aprosio, Fondazione Bruno Kessler (ITALY)
Marco Polignano, University of Bari (ITALY)
Paolo Rosso, Universitat Politècnica de València (SPAIN)
Manuela Sanguinetti, University of Cagliari (ITALY)
Efstathios Stamatatos, University of Aegean (GREECE)
Arkaitz Zubiaga, Queen Mary University of London (UNITED KINGDOM)

- **Keynote Speakers:**

Walter Daelemans - Universiteit Antwerpen (BELGIUM)
Viviana Patti - University of Turin (ITALY)

Table of Contents

<i>A First Attempt at Unreliable News Detection in Swedish</i> Ricardo Muñoz Sánchez, Eric Johansson, Shakila Tayefeh and Shreyash Kad	1
<i>BanglaHateBERT: BERT for Abusive Language Detection in Bengali</i> Md Saroar Jahan, Mainul Haque, Nabil Arhab and Mourad Oussalah	8
<i>A Comparison of Machine Learning Techniques for Turkish Profanity Detection</i> Levent Soykan, Cihan Karsak, ilknur Durgar Elkahlout and Burak Aytan	16
<i>Features and Categories of Hyperbole in Cyberbullying Discourse on Social Media</i> Simona Ignat and Carl Vogel	25

Conference Program

A First Attempt at Unreliable News Detection in Swedish

Ricardo Muñoz Sánchez, Eric Johansson, Shakila Tayefeh and Shreyash Kad

BanglaHateBERT: BERT for Abusive Language Detection in Bengali

Md Saroar Jahan, Mainul Haque, Nabil Arhab and Mourad Oussalah

A Comparison of Machine Learning Techniques for Turkish Profanity Detection

Levent Soykan, Cihan Karsak, ilknur Durgar Elkahlout and Burak Aytan

Features and Categories of Hyperbole in Cyberbullying Discourse on Social Media

Simona Ignat and Carl Vogel