

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

**Resources and Techniques for
User and Author Profiling in Abusive Language
(ResT-UP 2020)**

PROCEEDINGS

Editors:

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

**Proceedings of the LREC 2020 workshop on
Resources and Techniques for
User and Author Profiling in Abusive Language
(ResT-UP 2020)**

Edited by:

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

ISBN: 979-10-95546-49-8

EAN: 9791095546498

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 LREC2020 Workshop on Resources and Techniques for User and Author Profiling in Abusive Language (ResT-UP).

This volume documents the Proceedings of the 1st Workshop on Resources and Techniques for User and Author Profiling in Abusive Language (ResT-UP), held online on 12 May 2020 as part of the LREC 2020 conference (International Conference on Language Resources and Evaluation).

The workshop aimed 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 the existing LRs, e.g., datasets, corpora, lexicons, and sharing ideas on stylometry techniques and features needed for profile information extraction and classification. ResT-UP targeted Profiling scholars and research groups, experts in Statistic and Stylistic Analysis of texts as well as computational linguists who investigate author profile 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 represented an opportunity to share profiling experiments with the scientific community and to show automatic detection techniques of abusive language on the Web. Despite the cancellation of LREC 2020 due to the COVID-19 international emergency, ResT-UP was organized online on Microsoft Teams on May 12th 2020 and the programme included three oral presentations, and featured an invited talk by Paolo Rosso. ResT-UP was attended by about fifty representatives of academic and industrial organisations. We would like to thank the invited speaker, all authors who contributed papers to this workshop edition and the Programme Committee members who provided valuable feedback during the review process.

Johanna Monti – L’Orientale University of Naples – UNIOR NLP Research Group

Valerio Basile – University of Turin

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
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)
Tommaso Caselli, University of Groningen (NETHERLANDS)
Walter Daelemans, University of Antwerp (BELGIUM)
Rossana Damiano, University of Turin (ITALY)
Maciej Eder, Pedagogical University of Kraków (POLAND)
Francesca Frontini, Université Paul Valéry Montpellier 3 (FRANCE)
Dimitrios Kokkinakis, University of Göteborg (SWEDEN)
Stefano Menini, Fondazione Bruno Kessler (ITALY)
Cataldo Musto, University of Bari (ITALY)
Malvina Nissim, University of Groningen (NETHERLANDS)
Michael Oakes, University of Wolverhampton (UNITED KINGDOM)
Alessio Palmero Aprosio, Fondazione Bruno Kessler (ITALY)
Viviana Patti, University of Turin (ITALY)
Marco Polignano, University of Bari (ITALY)
Paolo Rosso, Universitat Politècnica de València (SPAIN)
Manuela Sanguinetti, University of Turin (ITALY)
Efsthathios Stamatatos, University of Aegean (GREECE)
Natalia Viani, King’s College London (UNITED KINGDOM)
Marcos Zampieri, Rochester Institute of Technology (U.S.A.)

Invited Speaker:

Paolo Rosso, Universitat Politècnica de València

Prof. Paolo Rosso is Full Professor at the Universitat Politècnica de València, Spain. Prof. Paolo Rosso received his PhD in Computer Science at the Trinity College University of Dublin, Ireland, in 1999. He is member of the Natural Language Engineering Lab. at Pattern Recognition and Human Language Technologies (PRHLT) Research Center. His research focuses on Author Profiling in Social Media, Irony Detection and Opinion Mining, Deceptive Opinion Detection, Stance Detection, Fake News Detection, Hate Speech Detection, Mixed-script Text Analysis and Plagiarism and Social Copying Detection.

Table of Contents

<i>Profiling Bots, Fake News Spreaders and Haters</i>	
Paolo Rosso	1
<i>An Indian Language Social Media Collection for Hate and Offensive Speech</i>	
Anita Saroj and Sukomal Pal	2
<i>Profiling Italian Misogynist: An Empirical Study</i>	
Elisabetta Fersini, Debora Nozza and Giulia Boifava	9
<i>Lower Bias, Higher Density Abusive Language Datasets: A Recipe</i>	
Juliet van Rosendaal, Tommaso Caselli and Malvina Nissim	14

Conference Program

Tuesday 12 May 2020

- 17:30–18:00 *Profiling Bots, Fake News Spreaders and Haters*
Paolo Rosso
- 16:30–16:50 *An Indian Language Social Media Collection for Hate and Offensive Speech*
Anita Saroj and Sukomal Pal
- 16:50–17:10 *Profiling Italian Misogynist: An Empirical Study*
Elisabetta Fersini, Debora Nozza and Giulia Boifava
- 17:10–17:30 *Lower Bias, Higher Density Abusive Language Datasets: A Recipe*
Juliet van Rosendaal, Tommaso Caselli and Malvina Nissim