

EACL Hackashop on News Media Content Analysis and Automated Report Generation 2021

Online
19 April 2021

Editors:

**Hannu Toivonen
Michele Boggia**

ISBN: 978-1-7138-2960-7

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571



Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2021) by the Association for Computational Linguistics
All rights reserved.

Printed with permission by Curran Associates, Inc. (2021)

For permission requests, please contact the Association for Computational Linguistics
at the address below.

Association for Computational Linguistics
209 N. Eighth Street
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006
Fax: 1-570-476-0860

acl@aclweb.org

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2633
Email: curran@proceedings.com
Web: www.proceedings.com

Table of Contents

Peer-reviewed Workshop Papers

<i>Adversarial Training for News Stance Detection: Leveraging Signals from a Multi-Genre Corpus.</i> Costanza Conforti, Jakob Berndt, Marco Basaldella, Mohammad Taher Pilehvar, Chryssi Giannitsarou, Flavio Toxvaerd and Nigel Collier	1
<i>Related Named Entities Classification in the Economic-Financial Context</i> Daniel De Los Reyes, Allan Barcelos, Renata Vieira and Isabel Manssour	8
<i>BERT meets Shapley: Extending SHAP Explanations to Transformer-based Classifiers</i> Enja Kokalj, Blaž Škrlič, Nada Lavrač, Senja Pollak and Marko Robnik-Šikonja	16
<i>Extending Neural Keyword Extraction with TF-IDF tagset matching</i> Boshko Koloski, Senja Pollak, Blaž Škrlič and Matej Martinc	22
<i>Zero-shot Cross-lingual Content Filtering: Offensive Language and Hate Speech Detection</i> Andraž Pelicon, Ravi Shekhar, Matej Martinc, Blaž Škrlič, Matthew Purver and Senja Pollak. . . .	30
<i>Exploring Linguistically-Lightweight Keyword Extraction Techniques for Indexing News Articles in a Multilingual Set-up</i> Jakub Piskorski, Nicolas Stefanovitch, Guillaume Jacquet and Aldo Podavini	35
<i>No NLP Task Should be an Island: Multi-disciplinarity for Diversity in News Recommender Systems</i> Myrthe Reuver, Antske Fokkens and Suzan Verberne	45
<i>TeMoTopic: Temporal Mosaic Visualisation of Topic Distribution, Keywords, and Context</i> Shane Sheehan, Saturnino Luz and Masood Masoodian	56
<i>Using contextual and cross-lingual word embeddings to improve variety in template-based NLG for automated journalism</i> Miia Rämö and Leo Leppänen	62
<i>Aligning Estonian and Russian news industry keywords with the help of subtitle translations and an environmental thesaurus</i> Andraž Repar and Andrej Shumakov	71
<i>Exploring Neural Language Models via Analysis of Local and Global Self-Attention Spaces</i> Blaž Škrlič, Shane Sheehan, Nika Eržen, Marko Robnik-Šikonja, Saturnino Luz and Senja Pollak	76
<i>Comment Section Personalization: Algorithmic, Interface, and Interaction Design</i> Yixue Wang	84
<i>Unsupervised Approach to Multilingual User Comments Summarization</i> Aleš Žagar and Marko Robnik-Šikonja	89

News Media Resources

EMBEDDIA Tools, Datasets and Challenges: Resources and Hackathon Contributions

Senja Pollak, Marko Robnik-Šikonja, Matthew Purver, Michele Boggia, Ravi Shekhar, Marko Pranjic, Salla Salmela, Ivar Krustok, Tarmo Paju, Carl-Gustav Linden, Leo Leppänen, Elaine Zosa, Matej Ulčar, Linda Freienthal, Silver Traat, Luis Adrián Cabrera-Diego, Matej Martinc, Nada Lavrač, Blaž Škrlj, Martin Žnidaršič, Andraž Pelicon, Boshko Koloski, Vid Podpečan, Janez Kranjc, Shane Sheehan, Emanuela Boros, Jose G. Moreno, Antoine Doucet and Hannu Toivonen 99

Hackathon Reports

A COVID-19 news coverage mood map of Europe

Frankie Robertson, Jarkko Lagus and Kaisla Kajava 110

Interesting cross-border news discovery using cross-lingual article linking and document similarity

Boshko Koloski, Elaine Zosa, Timen Stepišnik-Perdih, Blaž Škrlj, Tarmo Paju and Senja Pollak 116

EMBEDDIA hackathon report: Automatic sentiment and viewpoint analysis of Slovenian news corpus on the topic of LGBTIQ+

Matej Martinc, Nina Perger, Andraž Pelicon, Matej Ulčar, Andreja Vezovnik and Senja Pollak 121

To Block or not to Block: Experiments with Machine Learning for News Comment Moderation

Damir Korencic, Ipek Baris, Eugenia Fernandez, Katarina Leuschel and Eva Salido 127

Implementing Evaluation Metrics Based on Theories of Democracy in News Comment Recommendation

Myrthe Reuver and Nicolas Mattis 134