EACL 2021
The 16th conference of the European Chapter
of the Association for Computational Linguistics

EACL Hackashop on News Media Content Analysis and Automated Report Generation

Proceedings

Hannu Toivonen and Michele Boggia, Editors

April 19, 2021

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-954085-13-8

The hackashop has been organized by the EMBEDDIA project (Cross-Lingual Embeddings for Less-Represented Languages in European News Media) with support from the European Union's Horizon 2020 research and innovation program under grant 825153.

Preface

Automated content analysis of news media, including both news articles and users' comments on them, can provide unparalleled insight into current events, interests and opinions, as well as trends and changes in them. The needs are varied, from the readers who consume news of their personal interest to journalists who keep track of what is going on in the world, try to understand what their readers think of various topics, or want to automate routine reporting.

The aim of Hackashop 2021 is to foster discussion and research on the combination of language technology and news media content. The hackashop provides a forum for both discussing scientific advances in analysis of news stories and their reader comments and in automated generation of reports, as well as for experimental work on identifying interesting phenomena in reader comments and reporting on them.

Accordingly, the hackashop was implemented in a dual format. A traditional track consisted of submission of scientific papers, their reviews and finally paper presentations. It was complemented by an active, experimentation-based track consisting of an online hackathon preceding the workshop, with presentation of the results in the joint workshop event. Both tracks shared the same topic, news media analysis and generation, and participants to the two tracks had a good amount of overlap.

In the workshop track, we encouraged submissions of long and short papers. Based on three experts reviews for each submission, weighing the contributions of the submission against its length, 13 papers were selected for presentation in the workshop event.

The online hackathon was organized during a three-week period in February 2021, with six participating teams. The challenges they addressed covered a broad range, as each team had the freedom to define their own aims. In the spirit of providing a joint forum for discussing both scientific advances and experimental work, five hackathon teams submitted short reports to be included in this proceedings.

We also include in this proceedings an overview paper on all the tools, models, datasets and challenges collected and provided for the hackathon, as a resource for future scientific and empirical work in the area of news media content analysis and automated report generation.

We were very happy to see several cross-disciplinary and cross-sector collaborations involving, e.g., computer scientists, social scientists and media industry, both in workshop papers and hackathon contributions. We were also happy to have numerous contributions that address multilingual settings and low-resource languages.

The workshop event on 19 April 2021 brings both tracks together, with presentations of both scientific workshop papers and empirical hackathon reports.

We would like to thank all workshop paper authors and hackathon participants for their contributions to the hackashop! We are thankful to the programme committee members for their insightful reviews of the workshop papers. We are equally thankful to the large number of experts who made tools, models, data and challenges available for the hackathon and provided support for the participants.

We are grateful to EACL for giving the opportunity to organize the hackashop with them and to experiment with a novel format. The organization was supported by the European Union's Horizon 2020 research and innovation program under grant 825153 (EMBEDDIA).

Organizing committee

Organizing Committee

- Hannu Toivonen (University of Helsinki, Finland), Chair
- Matthew Purver (Queen Mary University of London, UK)
- Senja Pollak (Jozef Stefan Institute, Slovenia)
- Nada Lavrač (Jozef Stefan Institute, Slovenia)
- Marko Robnik-Šikonja (University of Ljubljana, Slovenia)
- Michele Boggia (University of Helsinki, Finland)
- Carl-Gustav Linden (University of Bergen, Norway)

Workshop Programme Committee

- Emanuela Boros (University of La Rochelle, France)
- Zoran Bosnić (University of Ljubljana, Slovenia)
- Hilde van den Bulck (Drexel University, USA)
- Nicholas Diakopoulos (Northwestern University, USA)
- Antoine Doucet (University of La Rochelle, France)
- Mark Granroth-Wilding (University of Helsinki, Finland)
- Adam Jatowt (Kyoto University, Japan)
- Maria Liakata (Queen Mary University of London, UK)
- Saturnino Luz (University of Edinburgh, UK)
- Matej Martinc (Jozef Stefan Institute, Slovenia)
- Marko Milosavljević (University of Ljubljana, Slovenia)
- Jose Moreno (IRIT, France)
- Kiem Hieu Nguyen (Hanoi university of science and technology, Vietnam)
- Lidia Pivovarova (University of Helsinki, Finland)
- Matej Ulčar (University of Ljubljana, Slovenia)
- Renata Vieira (University of Evora, Portugal)
- Carl Vogel (Trinity College Dublin, Ireland)
- Ivan Vulić (University of Cambridge, UK)
- Slavko Žitnik (University of Ljubljana, Slovenia)

Hackathon Experts

- Emanuela Boros (University of La Rochelle)
- Luis Adrián Cabrera-Diego (University of La Rochelle)
- Linda Freienthal (TEXTA OÜ)
- Boshko Koloski (Jožef Stefan Institute)
- Janez Kranjc (Jožef Stefan Institute)
- Ivar Krustok (Ekspress Meedia)
- Leo Leppänen (University of Helsinki)
- Matej Martinc (Jožef Stefan Institute)
- Jose G. Moreno (University of Toulouse)
- Tarmo Paju (Ekspress Meedia)
- Andraž Pelicon (Jožef Stefan Institute)
- Vid Podpečan (Jožef Stefan Institute)
- Marko Pranjić (Trikoder d.o.o.)
- Salla Salmela (Suomen Tietotoimisto STT)
- Shane Sheehan (University of Edinburgh)
- Ravi Shekhar (Queen Mary University of London)
- Blaž Škrlj (Jožef Stefan Institute)
- Silver Traat (TEXTA OÜ)
- Matej Ulčar (University of Ljubljana)
- Martin Žnidaršič (Jožef Stefan Institute)
- Elaine Zosa (University of Helsinki)

Table of Contents

Peer-reviewed Workshop Papers

News Media Resources

Hackathon Reports