

EMNLP 2017

**Second workshop on
Natural Language Processing meets Journalism**

Proceedings of the Workshop

September 7, 2017
Copenhagen, Denmark

©2017 The Association for Computational Linguistics

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 978-1-945626-88-3

Introduction

During the years NLP has matured a suites of technologies able to cope with many problems raised by the contemporary need of global information, and so, it is the high time for the NLP to get engaged in the mass media process as an active partner for both journalists and readers. The EMNLP workshop - at its second edition after the IJCAI-2016 workshop held on July 10th, 2016 in New York - has the potential to attract the interest of researchers both in computational linguistics and journalism, and of professionals in the news production system. The main goal is to have a forum in which it will be possible to share and discuss advancements in natural language processing and real needs in the field of journalism. The workshop received an unexpected number of submissions and the program committee identified 19 papers of high quality. We are happy to notice that papers submitted to this second edition display a large variety of topics that we consider at the core of our initiative. Many interesting aspects of journalism in mass-media age, from discovering bias and incongruent headlines in news to analyzing the content of social media in order to capture hot topics and trends are present. We can confidently say that fundamental topics of natural language processing have been covered this year, a fact that makes all the effort of organizing this workshop worthwhile. There are four excellent papers running for the best paper award and we already know it is a close call. Some of the accepted papers are likely to stir vivid debates, which, as the last year experience proves it, will turn into papers next year. In fact, we can see that some of the ideas that emerged from the papers and the discussions we carried out in the first edition are properly represented in this second edition. While we cannot partition the papers into three or four well defined classes, we can see that the NLP researchers have understood and have developed technologies able to cope with bias detection, sentiment analysis and relevance of specific information.

In an era in which there is no single point of view and the "objectiveness of information" lacks a unilateral standard, playing with language register and clarity in a piece of news are powerful instruments at journalist disposal. NLP may raise some flags on details, and papers such as "Tracking Bias in News Sources Using Social Media: the Russia-Ukraine Maidan Crisis of 2013–2014", "Incongruent Headlines: Yet Another Way to Mislead Your Readers", "Fake news stance detection using stacked ensemble of classifiers", "From Clickbait to Fake News Detection: An Approach based on Detecting the Stance of Headlines to Articles", "An NLP Analysis of Exaggerated Claims in Science News" present interesting and efficient ways to warn the reader against possible abuses in the media.

Sentiment analysis is an important component in understanding the impact of a piece of news and few papers at this workshop focus on this aspect: "Comparing Attitudes to Climate Change in the Media using sentiment analysis based on Latent Dirichlet Allocation", "Improved Abusive Comment Moderation with User Embeddings", "Deception Detection in News Reports in the Russian Language: Lexics and Discourse", "Using New York Times Picks to Identify Constructive Comments", "Using New York Times Picks to Identify Constructive Comments", "Predicting News Values from Headline Text and Emotions".

A group of papers deals with a crucial issue of modern journalism, i.e., how to manage the enormous flow of available information: "What to Write? A topic recommender for journalists", "Data-to-text generation for tailored soccer journalism", "Language-based Construction of Explorable News Graphs for Journalists", "Storyteller: Visual Analytics of Perspectives on Rich Text Interpretations", "Analyzing the Revision Logs of a Japanese Newspaper for Article Quality Assessment", "Unsupervised Event Clustering and Aggregation from Newswire and Web Article", "Semantic Storytelling, Cross-lingual Event Detection and other Semantic Services for a Newsroom Content Curation Dashboard", "A News Chain Evaluation Methodology along with a Lattice-based Approach for News Chain Construction".

In conclusion, we are very happy and proud about the high quality papers dealing with essential topics in modern journalism. We believe that more and more NLP will have an impact on journalism, the force that truly and directly represents all us in the confrontation with the social problems.

Octavian Popescu and Carlo Strapparava

Organizers:

Octavian Popescu, IBM Watson Research Center, USA
Carlo Strapparava, FBK-irst, Italy

Program Committee:

Enrique Alfonseca, Google
Tommaso Caselli, VU Amsterdam University
Dan Cristea, University of Iasi
Liviu P. Dinu, University of Bucharest
Elena Erdmann, TU Dortmund University
Song Feng, IBM Research
Lorenzo Gatti, FBK-irst
Marco Guerini, FBK-irst
James Hamilton, Stanford University
Mark Hansen, Columbia university
Orin Hargraves, University of Colorado
Daisuke Kawahara, Kyoto University
Kristian Kersting, TU Dortmund University
Shervin Malmasi, Harvard Medical School
Rada Mihalcea, University of Michigan
Preslav Nakov, Qatar Computing Research Institute, HBKU
Vivi Nastase, University of Heidelberg
Gözde Özbal, FBK-irst
Martha Palmer, University of Colorado
Siddharth Patwardhan, IBM Research TJ Watson
Preethi Raghavan, IBM Research TJ Watson
Mattia Rigotti IBM Research TJ Watson
Paolo Rosso, Universitat Politècnica de València
Amanda Stent, Bloomberg
Serra Sinem Tekiroglu, FBK-irst
Ngoc Phuoc An Vo, IBM Research
Marcos Zampieri, University of Cologne
Torsten Zesch, University of Duisburg-Essen

Table of Contents

<i>Predicting News Values from Headline Text and Emotions</i> Maria Pia di Buono, Jan Šnajder, Bojana Dalbelo Basic, Goran Glavaš, Martin Tutek and Natasa Milic-Frayling	1
<i>Predicting User Views in Online News</i> Daniel Hardt and Owen Rambow	7
<i>Tracking Bias in News Sources Using Social Media: the Russia-Ukraine Maidan Crisis of 2013–2014</i> Peter Potash, Alexey Romanov, Mikhail Gronas, Anna Rumshisky and Mikhail Gronas	13
<i>What to Write? A topic recommender for journalists</i> Alessandro Cucchiarelli, Christian Morbidoni, Giovanni Stilo and Paola Velardi	19
<i>Comparing Attitudes to Climate Change in the Media using sentiment analysis based on Latent Dirichlet Allocation</i> Ye Jiang, Xingyi Song, Jackie Harrison, Shaun Quegan and Diana Maynard	25
<i>Language-based Construction of Explorable News Graphs for Journalists</i> Rémi Bois, Guillaume Gravier, Eric Jamet, Emmanuel Morin, Pascale Sébillot and Maxime Robert	31
<i>Storyteller: Visual Analytics of Perspectives on Rich Text Interpretations</i> Maarten van Meersbergen, Piek Vossen, Janneke van der Zwaan, Antske Fokkens, Willem van Hage, Inger Leemans and Isa Maks	37
<i>Analyzing the Revision Logs of a Japanese Newspaper for Article Quality Assessment</i> Hideaki Tamori, Yuta Hitomi, Naoaki Okazaki and Kentaro Inui	46
<i>Improved Abusive Comment Moderation with User Embeddings</i> John Pavlopoulos, Prodromos Malakasiotis, Juli Bakagianni and Ion Androutsopoulos	51
<i>Incongruent Headlines: Yet Another Way to Mislead Your Readers</i> Sophie Chesney, Maria Liakata, Massimo Poesio and Matthew Purver	56
<i>Unsupervised Event Clustering and Aggregation from Newswire and Web Articles</i> Sven Ribeiro, Olivier Ferret and Xavier Tannier	62
<i>Semantic Storytelling, Cross-lingual Event Detection and other Semantic Services for a Newsroom Content Curation Dashboard</i> Julian Moreno-Schneider, Ankit Srivastava, Peter Bourgonje, David Wabnitz and Georg Rehm .	68
<i>Deception Detection in News Reports in the Russian Language: Lexics and Discourse</i> Dina Pisarevskaya	74
<i>Fake news stance detection using stacked ensemble of classifiers</i> James Thorne, Mingjie Chen, Giorgos Myriantous, Jiashu Pu, Xiaoxuan Wang and Andreas Vlachos	80
<i>From Clickbait to Fake News Detection: An Approach based on Detecting the Stance of Headlines to Articles</i> Peter Bourgonje, Julian Moreno Schneider and Georg Rehm	84

<i>'Fighting' or 'Conflict'? An Approach to Revealing Concepts of Terms in Political Discourse</i>	
Linyuan Tang and Kyo Kageura	90
<i>A News Chain Evaluation Methodology along with a Lattice-based Approach for News Chain Construction</i>	
Mustafa Toprak, Özer Özkahraman and Selma Tekir	95
<i>Using New York Times Picks to Identify Constructive Comments</i>	
Varada Kolhatkar and Maite Taboada	100
<i>An NLP Analysis of Exaggerated Claims in Science News</i>	
YINGYA LI, Jieke Zhang and Bei Yu	106

Conference Program

September 7th

Morning

Oral Presentations

Predicting News Values from Headline Text and Emotions

Maria Pia di Buono, Jan Šnajder, Bojana Dalbelo Basic, Goran Glavaš, Martin Tutek and Natasa Milic-Frayling

Predicting User Views in Online News

Daniel Hardt and Owen Rambow

Tracking Bias in News Sources Using Social Media: the Russia-Ukraine Maidan Crisis of 2013–2014

Peter Potash, Alexey Romanov, Mikhail Gronas, Anna Rumshisky and Mikhail Gronas

What to Write? A topic recommender for journalists

Alessandro Cucchiarelli, Christian Morbidoni, Giovanni Stilo and Paola Velardi

Comparing Attitudes to Climate Change in the Media using sentiment analysis based on Latent Dirichlet Allocation

Ye Jiang, Xingyi Song, Jackie Harrison, Shaun Quegan and Diana Maynard

Language-based Construction of Explorable News Graphs for Journalists

Rémi Bois, Guillaume Gravier, Eric Jamet, Emmanuel Morin, Pascale Sébillot and Maxime Robert

Storyteller: Visual Analytics of Perspectives on Rich Text Interpretations

Maarten van Meersbergen, Piek Vossen, Janneke van der Zwaan, Antske Fokkens, Willem van Hage, Inger Leemans and Isa Maks

Analyzing the Revision Logs of a Japanese Newspaper for Article Quality Assessment

Hideaki Tamori, Yuta Hitomi, Naoaki Okazaki and Kentaro Inui

Improved Abusive Comment Moderation with User Embeddings

John Pavlopoulos, Prodromos Malakasiotis, Juli Bakagianni and Ion Androutsopoulos

September 7th (continued)

Lunch

Invited Talk

Cristian Danescu-Niculescu-Mizil - Cornell University

Poster Presentations

Incongruent Headlines: Yet Another Way to Mislead Your Readers

Sophie Chesney, Maria Liakata, Massimo Poesio and Matthew Purver

Unsupervised Event Clustering and Aggregation from Newswire and Web Articles

Swen Ribeiro, Olivier Ferret and Xavier Tannier

Semantic Storytelling, Cross-lingual Event Detection and other Semantic Services for a Newsroom Content Curation Dashboard

Julian Moreno-Schneider, Ankit Srivastava, Peter Bourgonje, David Wabnitz and Georg Rehm

Deception Detection in News Reports in the Russian Language: Lexics and Discourse

Dina Pisarevskaya

Fake news stance detection using stacked ensemble of classifiers

James Thorne, Mingjie Chen, Giorgos Myriantous, Jiashu Pu, Xiaoxuan Wang and Andreas Vlachos

From Clickbait to Fake News Detection: An Approach based on Detecting the Stance of Headlines to Articles

Peter Bourgonje, Julian Moreno Schneider and Georg Rehm

'Fighting' or 'Conflict'? An Approach to Revealing Concepts of Terms in Political Discourse

Linyuan Tang and Kyo Kageura

A News Chain Evaluation Methodology along with a Lattice-based Approach for News Chain Construction

Mustafa Toprak, Özer Özkahraman and Selma Tekir

Using New York Times Picks to Identify Constructive Comments

Varada Kolhatkar and Maite Taboada

September 7th (continued)

An NLP Analysis of Exaggerated Claims in Science News
YINGYA LI, Jieke Zhang and Bei Yu

Best paper announcement and Conclusions

