

The Ants Have Megaphones Now: Text Mining and Summarization for News and Social Media Streams

Georgiana Ifrim
School of Computer Science
University College Dublin, Ireland

Abstract

In this talk we present some of our research on employing news and social media streams as complementary means to tell and understand a news story. In particular, we focus on the techniques, tools and datasets we developed for real-time social tag annotations of news, and how we employed Twitter hashtags as a novel type of metadata to facilitate news indexing, search, story tracking and summarization. We discuss the importance of developing more effective ways of condensing information for human consumption, in particular through methods that automatically generate structured summaries of news stories. We also present some benchmarks and evaluation studies for structured summarization of news and point out remaining challenges and gaps in the evaluation of structured summaries. Most of our tools and datasets are available online from <https://github.com/mlgig>.

Short Bio

Dr. Georgiana Ifrim is an Associate Professor at the School of Computer Science, University College Dublin (UCD), Ireland, Co-Lead of the SFI Centre for Research Training in Machine Learning (ML-Labs) and SFI Funded Investigator with the Insight SFI Research Centre for Data Analytics and the VistaMilk SFI Research Centre. She is Director of Graduate Research at the School of Computer Science, UCD.

Prior to this, she held research fellow and post-doctoral positions with the Insight Centre for Data Analytics, University College Dublin, Ireland, Cork Constraint Computation Centre, University College Cork, Ireland, and Bioinformatics Research Centre, Aarhus University, Denmark. She holds a PhD and MSc from Max-Planck Institute for Informatics, Germany, and a BSc from University of Bucharest, Romania. Dr. Ifrim's research focuses on developing scalable predictive models for machine learning and data mining applications. She has developed new methods for sequence learning, time series classification, text mining and real-time prediction for news and social streams. Dr. Ifrim has worked in application domains ranging from Web mining, news and social media, energy, biology and sports science. Her current research focuses on the design of efficient and interpretable learning models for sequences (e.g., DNA, time series), and on text mining and summarization for news and social media.

Acknowledgements

This work was funded by Science Foundation Ireland (SFI) under grant number 12/RC/2289 and 12/RC/2289_P2, as well as the Irish Research Council (IRC) under grant number EBPPG/2018/237.

Copyright © 2020 by paper's authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

In: A. Jorge, R. Campos, A. Jatowt, A. Aizawa (eds.): Proceedings of the first AI4Narratives Workshop, Yokohama, Japan, January 2021, published at <http://ceur-ws.org>