

ECIR 2018: Text2Story Workshop - Narrative Extraction from Texts

Alípio Jorge^{1,2}, Ricardo Campos^{1,3}, Adam Jatowt⁴, Sérgio Nunes^{1,2}
Conceição Rocha¹, João Paulo Cordeiro^{1,5}
Arian Pasquali^{1,2} and Vitor Mangaravite¹

¹INESC TEC, LIAAD and CSIG, Porto, Portugal

²University of Porto - FCUP and FEUP, Porto, Portugal

³Polytechnic Institute of Tomar – Smart Cities Research Center, Tomar, Portugal

⁴Kyoto University, Kyoto, Japan

⁵University of Beira Interior, Covilhã, Portugal

amjorge@fc.up.pt, ricardo.campos@ipt.pt, adam@dl.kuis.kyoto-u.ac.jp, sergio.nunes@fe.up.pt

Abstract

The 1st International Workshop on Narrative Extraction from Texts (Text2Story 2018) was held in conjunction with the 40th European Conference on Information Retrieval, ECIR 2018, Grenoble on the 26th March 2018. The workshop aimed to help foster the collaboration of researchers on a wide range of multidisciplinary issues related to the text-to-narrative-structure. The program consisted of two keynote talks, six research presentations, a poster session and a slot for demo presentations. This report briefly summarizes the workshop. More information about the workshop is available at <http://text2story18.inesctec.pt>

1 Introduction

The increasing availability of text information in the form of news articles, comments or posts poses new challenges for those who aim to understand the storyline of an event. Although understanding natural language text has improved over the last couple of years with research works emerging on the grounds of information extraction and text mining, the problem of constructing consistent narrative structures is yet to be solved. We still have a challenging path ahead of us for the development and improvement of algorithms that automatically identify, interpret and relate the different elements of a narrative, either in a single text or spread among different sources. The aim of this workshop was to foster the discussion of recent advances in the intersection between Information Retrieval (IR) and formal narrative representations from texts. More specifically, it was our intention to capture a wide range of multidisciplinary issues related to the text-to-narrative-structure and to its various related tasks. The workshop featured a diversity of tasks and techniques with promising results.

We had four full papers and two short papers selected by the programme committee from a total of 11 submissions. In addition to the selected papers, the workshop featured two invited speakers. Udo Kruschwitz on “Users2Story - On the Importance of Understanding Searchers’ Information Needs”, and Eric Gaussier, who addressed “Word embeddings, information retrieval and textual entailment”.

A special issue of the Information Processing & Management (IPM) Journal is planned on the topics of the workshop.

The workshop chairs were Alípio Jorge, Ricardo Campos, Adam Jatowt and Sérgio Nunes. The proceedings were setup by our Proceedings Chairs Conceição Rocha and João Paulo Cordeiro and are available at CEUR (<http://ceur-ws.org/Vol-2077/>). Arian Pasquali was our web chair (<http://text2story18.inesctec.pt>) and Vitor Mangaravite was responsible for workshop information dissemination.

2 Keynote by Udo Kruschwitz

The program started with a keynote talk given by Udo Kruschwitz from University of Essex focused on uncovering the storyline of events. On this keynote talk Udo discussed the various search and browsing patterns of a user on the web and on a corporate information retrieval system, which could give us a better understanding of the user's actual information needs. Log files were also discussed as a useful resource to build user profiles.

3 Keynote by Eric Gaussier

The other talk was given on the afternoon session of the workshop by Eric Gaussier from the University of Grenoble Alps. The talk focused on one of the currently preferred representations for words in various NLP and IR tasks. In this presentation, Eric reviewed the main embedding approaches used in IR and explained to which extent they lead to improved IR performance. The possibility to extend current word embeddings with syntactic information and the impact of doing so in the context of different NLP tasks were also discussed.

4 Paper Presentations

We had 6 research presentations by Venumadhav Kattagoni, Júlia Galántai, Attila Gulyás, Muhammad Nihal Hussain, Jason J. Jung and Jorge Valverde-Rebaza. The accepted papers have the following titles and authors and can be found at <http://ceur-ws.org/Vol-2077> :

- "IREvent2Story: A Novel Mediation Ontology and Narrative Generation" by Venumadhav Kattagoni and Navjyoti Singh;
- "Gossip is More than Just Story Telling Topic Modelling and Quantitative Analysis on a Spontaneous Speech Corpus" by Boróka Pápay, Bálint Kubik and Júlia Galántai;
- "Text Network Analysis and Visualization of Hungarian, Communist-era Political Reports" by Attila Gulyás, Martina K. Szabó, István Boros Jr. and Gergő Havadi;
- "Analyzing Shift in Narratives Regarding Migrants in Europe via Blogosphere" by Muhammad Nihal Hussain, Kiran Kumar Bandeli, Samer Al-Khateeb, Nitin Agarwal;
- "Measuring Character-based Story Similarity by Analyzing Movie Scripts" by O-Joun Lee, Nayoung Jo and Jason J. Jung;
- "Job Recommendation Based on Job Seeker Skills: An Empirical Study" by Jorge Valverde-Rebaza, Ricardo Puma, Paul Bustios and Nathalia C. Silva.

5 Poster Session

In our poster session presenters of the papers and attendees of the workshop had the opportunity to discuss their research and interact in a more informal way, networking and socializing. This was a very dynamic session.

6 Demo Session

The workshop ended with two demo presentations from Adam Jatowt and Vitor Mangaravite. The first one was designed for discovering the story of word semantic evolution based on Google Books data. Vitor presented the YAKE!¹ keyword extractor, a new approach to extract keywords from single documents which was awarded the Best Short Paper of the ECIR'18 conference.

7 Programme Committee

The following researchers and industry experts have served on the Programme Committee of Text2Story 2018:

- Klaus Berberich (Max Planck Institute for Informatics)
- Marc Spaniol (Normandie University)
- Miguel Martinez-Alvarez (Signal)
- Dhruv Gupta (Max Planck Institute for Informatics)
- Federico Nanni (University of Mannheim)
- Yihong Zhang (Kyoto University)
- Udo Kruschwitz (University of Essex)
- Grace Hui Yang (Georgetown University)
- Gerasimos Lampouras (The University of Sheffield)
- Denilson Barbosa (University of Alberta)
- Sumit Bhatia (IBM)
- Yating Zhang (Kyoto University)
- Nina Tahmasebi (University of Gothenburg)
- Jaime Arguello (University of North Carolina at Chapel Hill)
- Akiko Aizawa (NII)
- António Horta Branco (University of Lisbon)
- Bruno Martins (University of Lisbon)
- Gael Dias (University of Caen)
- Henrique Lopes Cardoso (University of Porto)
- Pablo Gamallo (University of Santiago de Compostela)
- Conceição Rocha (INESC TEC)
- João Paulo Cordeiro (Universidade da Beira Interior)
- Fernando Batista (INESC-ID & ISCTE-IUL)
- Álvaro Figueira (INESC TEC & University of Porto)
- Nuno Moniz (INESC TEC)
- Paulo Quaresma (University of Evora)
- Sebastião Pais (Universidade da Beira Interior)
- Pedro Saleiro (University of Chicago)

Acknowledgements

We would like to thank ECIR2018 organizers for helping us to organize the workshop, our keynote speakers, the program committee, the paper authors, and the participants. This work is partially funded by the ERDF through the COMPETE 2020 Programme within project POCI-01-0145-FEDER-006961, and by National Funds through the FCT as part of project UID/EEA/50014/2013 and of project UID/MAT/00212/2013. It was also financed by MIC SCOPE (171507010) and by Project “TEC4Growth - Pervasive Intelligence, Enhancers and Proofs of Concept with Industrial Impact/NORTE-01-0145-FEDER-000020” which is financed by the NORTE 2020, under the Portugal 2020, and through the ERDF.

¹ <http://bit.ly/YakeDemoECIR2018>