

# Report on the Third International Workshop on Narrative Extraction from Texts (Text2Story 2020)

Ricardo Campos  
INESC TEC, Portugal  
Ci2 - Smart Cities Research Center, Polytechnic Institute of Tomar, Portugal  
*ricardo.campos@ipt.pt*

Alípio M. Jorge  
INESC TEC, Portugal  
FCUP - University of Porto, Portugal

Adam Jatowt  
Kyoto University, Japan

Sumit Bhatia  
IBM Research AI, India

Arian Pasquali  
INESC TEC, Portugal

João Paulo Cordeiro  
INESC TEC, Portugal  
University of Beira Interior, Portugal

Conceição Rocha  
INESC TEC, Portugal

Behrooz Mansouri  
Rochester Institute of Technology,  
Rochester, NY, USA

Brenda Santana  
INESC TEC, Portugal  
FCUP - University of Porto, Portugal

## Prologue

The Third International Workshop on Narrative Extraction from Texts (Text2Story'20 [<https://text2story20.inesctec.pt/>]) was held on the 14th of April 2020, in conjunction with the 42<sup>nd</sup> European Conference on Information Retrieval (ECIR 2020). This year due to the Covid-19 outbreak the Text2Story workshop was held online on Zoom platform. During the course of the day, an average of more than 60 attendees had the opportunity to follow-up and discuss the recent advances in extraction and formal representation of narratives. The workshop consisted of two invited keynotes and thirteen paper presentations. The proceedings of the workshop are available online at <http://ceur-ws.org/Vol-2593/>

---

# 1 Background

The Text2Story series Workshops is aimed at bringing together researchers from diverse, related fields such as IR, NLP, AI, design and visualization to come up and share the recent advances in their respective fields towards narrative understanding. Building upon the success of the past two workshops [Jorge et al., 2018, 2019a] and on the Text2Story Special Issue at IPM Journal [Jorge et al., 2019b], this year, we organized the third edition, held online under the umbrella of ECIR 2020. This report, summarizes the key activities at the workshop. The Program Chairs were Ricardo Campos, Alípio Jorge, Adam Jatowt and Sumit Bhatia. The proceedings were setup by our Proceedings Chairs João Paulo Cordeiro and Conceição Rocha. Arian Pasquali was our Web Chair, Behrooz Mansouri the dissemination chair and Brenda Santana our research volunteer during the workshop event. The following directs the interested reader for the online proceedings, presentations and recordings of the talks given at the workshop.

**YouTube:** [https://www.youtube.com/channel/UC8ysv2TEpQZz\\_-v5T5nZSSA/videos](https://www.youtube.com/channel/UC8ysv2TEpQZz_-v5T5nZSSA/videos)

**Presentation Slides:** [https://text2story20.inesctec.pt/text2story2020\\_slides.zip](https://text2story20.inesctec.pt/text2story2020_slides.zip)

**Online Proceedings:** <http://ceur-ws.org/Vol-2593/>

## 2 The Plot

The workshop program consisted of two invited keynote talks, ten regular research papers, two position papers and one demonstration. The papers presented at the workshop covered diverse aspects of the narrative extraction problem ranging from applications of deep learning to narrative extraction, sentiment analysis, event-correference resolution to applications in different domains such as legal text, news and social media.

### 2.1 Keynotes

**Sebastião Miranda**, head of development at Priberam<sup>1</sup>, talked about monitoring multi-lingual, multi-media sources of information to discover relevant stories, events, and topics of interest to a specific user. He presented the solutions developed at Priberam for this purpose by training text retrieval models and integrating user feedback in the media monitoring pipeline. **Mark Finlayson**, assistant professor at School of Computing and Information Sciences, Florida - US<sup>2</sup>, presented an overview of recent advances in Natural Language Processing applied to the task of narrative analysis in text. The techniques covered in his talk were focused on four major tasks for narrative analysis: (1) story detection; (2) character detection; (3) event detection; and (4) timeline extraction.

### 2.2 Narrative Detection and Extraction

Kanjirangat et al. [2020] presented their work on utilizing deep learning based approaches for narrative text understanding. They proposed a transformer based classifier to capture static

---

<sup>1</sup><https://www.priberam.com/>

<sup>2</sup><https://www.fiu.edu/>

---

relations between different characters of a narrative and dynamic word embeddings to reflect the temporal evolution of relations between the different characters. Lee and Kim [2020] presented their work on assessing the fluency of a narrative extracted from textual data. They considered two aspects of fluency in narratives – *consistency*, in terms of topics and entities covered; and *rapidity*, in terms of the pace at which the narrative changes. Lyu et al. [2020] described how contextual information and external knowledge can be utilized to improve sentiment analysis of narrative text.

Effective narrative extraction from texts require identification and coreference resolution of events mentioned in different documents. Bugert et al. [2020] described an approach to annotate documents with cross-subtopics event coreference links in a scalable fashion and present a Football Coreference Corpus (FCC) created following their approach. The corpus consists of 451 sports news reports spanning over 15,000 sentences and covering 217 events.

## 2.3 Applications

Elsafoury [2020] presented a case study exploring the use of Twitter for identifying actions and events by authorities aimed at curbing protests during the Turkish Gezi Park protests in 2013. The authors also make available a crowd-sourced dataset to enable further research on this important problem.

Hausner et al. [2020] presented a framework to construct timelines for court cases by extracting key temporal phrases from court case documents. Such timelines can provide useful insights about the chronological evolution of court cases and different entities involved. Bandeli et al. [2020] presented a framework for narrative extraction in blogs where the text is diverse (created by individual contributors) and potentially noisy, due to lack of any editorial policies.

Berhe et al. [2020] discussed the problem of establishing links between different scenes in a series of multi-media collection (such as a TV series) for better organization of the content. For example, a story arc could be spread across multiple episodes or even different seasons. They present a dataset with hand annotated links between different scenes of the first two seasons of the TV series Game of Thrones. The dataset consists of 444 scenes involving 154 characters organized into 46 stories.

One major application of Narrative extraction from text is in analysis of news and event data as it can help the readers get an overview of an evolving event from multiple sources. To this end, Linger and Hajaiej [2020] described their system for clustering multi-lingual news articles and arranging them into coherent stories for easier consumption by the readers. Chakraborty et al. [2020] presented an approach for constructing a network of key characters involved in an event and relationships between them. Finally, Piskorski et al. [2020] demonstrated their tool for browsing and visualizing event-centric timelines from news articles. Given a target entity of interest, the system extracts information about events involving the entity from a stream of news articles.

## 2.4 Looking Ahead

The menace of fake news and misinformation has proliferated our digital lives and the techniques used for spreading such malicious information are becoming increasingly complex. Often, such articles present information that may be factually correct, but is presented using subtle tools such

---

as framing, selective reporting, and argument distortions to manipulate the reader. [Mensio et al. \[2020\]](#) proposed an integrated framework to identify such subtle clues and narrative signals in text and highlighted the differences in narratives as presented by different news sources.

Along similar lines, [Suchanek \[2020\]](#) posited that the current representation of *facts* about the World in Knowledge Graphs lacks the necessary details to handle complex tasks. He argued for the need to develop better information extraction tools to augment existing knowledge bases by incorporating richer meta-data such as relations of higher granularity, evidences and supporting facts for information stored in the knowledge bases and dependencies between different events. Access to such a rich knowledge base has the potential to improve fake news detection, controversy modeling, detecting frauds, and, in general, better understanding of text.

### 3 The Characters

Due to Covid-19 outbreak, the Text2Story workshop was run fully online this year through zoom platform (see Figure 1). This new realm forced a quick adaptation of the organization who had to fit a program to accommodate the different time-zones of presenters and session chairs. We owe the workshop success to the diverse set of participants, from academia and industry, who established a new record of more than 60 attendees (on average) during all day. The enthusiasm and active participation of the attendees, ensured that the virtual edition of the workshop was as intellectually stimulating as the past two physical editions.



Figure 1: Overview of the Text2Story2020 Online Workshop

---

The workshop program included a total of 13 research papers (out of 20 submissions) from 37 authors and two invited keynote talks. In addition to this, we had six session chairs who brilliantly moderated the presentation of the papers. Our thanks go to Arian Pasquali (INESC TEC), Ismail Sengor Altingovde (Middle East Technical University), Jeremy Pickens (OpenText), Marc Spaniol (Université de Caen Normandie), Paulo Quaresma (University of Évora) and Satya Almasian (Heidelberg University). Finally, we would like to acknowledge the effort and valuable contribution of the researchers and industry experts that have served on the Program Committee of the Text2Story'20 workshop. Our thanks go to:

- Álvaro Figueira (INESC TEC & University of Porto)
- Andreas Spitz (École polytechnique fédérale de Lausanne)
- António Horta Branco (University of Lisbon)
- Arian Pasquali (INESC TEC)
- Bruno Martins (IST and INESC-ID - Instituto Superior Técnico, University of Lisbon)
- Daniel Gomes (FCT/Arquivo.pt)
- Daniel Loureiro (University of Porto)
- Denilson Barbosa (University of Alberta)
- Dhruv Gupta (Max Planck Institute for Informatics)
- Dwaipayan Roy (ISI Kolkata, India)
- Dyaa Albakour (Signal)
- Gaël Dias (Normandie University)
- Henrique Lopes Cardoso (University of Porto)
- Ismail Sengor Altingovde (Middle East Technical University)
- Jeffery Ansah (BHP)
- Jeremy Pickens (OpenText)
- João Magalhães (Universidade Nova de Lisboa)
- Kiran Kumar Bandeli (Walmart Inc.)
- Ludovic Moncla (INSA Lyon)
- Marc Spaniol (Université de Caen Normandie)
- Mark Finlayson (Florida International University)

- 
- Mengdie Zhuang (The University of Sheffield)
  - Nina Tahmasebi (University of Gothenburg)
  - Nuno Moniz (LIAAD/INESC TEC)
  - Pablo Gamallo (University of Santiago de Compostela)
  - Paulo Quaresma (University of Évora)
  - Preslav Nakov (Qatar Computing Research Institute (QCRI))
  - Ross Purves (University of Zurich)
  - Satya Almasian (Heidelberg University)
  - Sebastiao Miranda (Priberam)
  - Sérgio Nunes (INESC TEC & University of Porto)
  - Udo Kruschwitz (University of Essex)
  - Vítor Mangaravite (UFMG)
  - Yihong Zhang (Kyoto University)

## 4 Applause

As part of an effort to promote top quality research, we decided to establish two prizes, the best paper award and the recognized reviewer award. The first is granted for the best paper published and presented at the Text2Story workshop. The latter, aims to recognize reviewers by their effort and the quality put on this noble task. The program chairs are pleased to announce that the Best Paper Award of the Text2Story 2020 Workshop went to Vani Kanjirangat, Simone Mellace and Alessandro Antonucci for their paper "Temporal Embeddings and Transformer Models for Narrative Text Understanding" [Kanjirangat et al., 2020]. In addition to this, the program chairs have recognized the following persons as recipients of the 2020 reviewer award for their insightful and valuable reviews - Andreas Spitz (École Polytechnique Fédérale de Lausanne), Bruno Martins (IST and INESC-ID - Instituto Superior Técnico, University of Lisbon), Ismail Sengor Altingovde (Middle East Technical University), Mark Finlayson (Florida International University) and Satya Almasian (Heidelberg University).

## 5 Epilogue

This was the third edition of the Text2Story workshop series. Our objective was to bring together the interested participants from different geographies and research expertise to collectively set the agenda for the emerging multi-disciplinary area of *narrative extraction from texts*. We have received increasingly positive response from the research community and the growing participation

---

in the workshops underlines the relevance of the topic. Thus, it is our aim to continue organizing the workshop series in IR and NLP venues to develop the field further.

### Acknowledgements

We would like to thank ECIR 2020 organizers for helping us to organize the workshop, our keynote speakers, members of the program committee, the paper authors, and the participants. The first two authors of this paper were financed by the ERDF – European Regional Development Fund through the North Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 and by National Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia within project PTDC/CCI-COM/31857/2017 (NORTE-01-0145-FEDER-03185)

## References

- Kiran Kumar Bandeli, Muhammad Nihal Hussain, and Nitin Agarwal. A framework towards computational narrative analysis on blogs. In *Proceedings of Text2Story - 3rd Workshop on Narrative Extraction From Texts, co-located with the 42nd European Conference on Information Retrieval, Text2Story@ECIR 2020, Lisbon, Portugal, April 14th*, pages 63–69, 2020.
- Aman Berhe, Camille Guinaudeau, and Claude Barras. Scene linking annotation and automatic scene characterization in tv series. In *Proceedings of Text2Story - 3rd Workshop on Narrative Extraction From Texts, co-located with the 42nd European Conference on Information Retrieval, Text2Story@ECIR 2020, Lisbon, Portugal, April 14th*, pages 47–53, 2020.
- Michael Bugert, Nils Reimers, Shany Barhom, Ido Dagan, and Iryna Gurevych. Breaking the subtopic barrier in cross-document event coreference resolution. In *Proceedings of Text2Story - 3rd Workshop on Narrative Extraction From Texts, co-located with the 42nd European Conference on Information Retrieval, Text2Story@ECIR 2020, Lisbon, Portugal, April 14th*, pages 23–29, 2020.
- Roshni Chakraborty, Srishti Bhandari, Nilotpall Chakraborty, and Ritwika Das. Eve2sign: Creating signed networks of news events. In *Proceedings of Text2Story - 3rd Workshop on Narrative Extraction From Texts, co-located with the 42nd European Conference on Information Retrieval, Text2Story@ECIR 2020, Lisbon, Portugal, April 14th*, pages 79–87, 2020.
- Fatma Elsafoury. Teargas, water cannons and twitter: A case study on detecting protest repression events in turkey 2013. In *Proceedings of Text2Story - 3rd Workshop on Narrative Extraction From Texts, co-located with the 42nd European Conference on Information Retrieval, Text2Story@ECIR 2020, Lisbon, Portugal, April 14th*, pages 5–13, 2020.
- Philip Hausner, Dennis Aumiller, and Michael Gertz. Time-centric exploration of court documents. In *Proceedings of Text2Story - 3rd Workshop on Narrative Extraction From Texts, co-located with the 42nd European Conference on Information Retrieval, Text2Story@ECIR 2020, Lisbon, Portugal, April 14th*, pages 31–37, 2020.

- 
- Alípio M. Jorge, Ricardo Campos, Adam Jatowt, and Sérgio Nunes. First international workshop on narrative extraction from text (text2story'18). In *Pasi G., Piwowarski B., Azzopardi L., Hanbury A. (eds). Advances in Information Retrieval. ECIR 2018 (Grenoble, France. March 26 – 29). Lecture Notes in Computer Science, vol 10772*, pages 833–834, 2018.
- Alípio M. Jorge, Ricardo Campos, Adam Jatowt, and Sumit Bhatia. Second international workshop on narrative extraction from text (text2story'19). In *Azzopardi L., Stein B., Fuhr N., Mayr P., Hauff C., Hiemstra D. (eds), Advances in Information Retrieval. ECIR'19 (Cologne, Germany. April 14 – 18). Lecture Notes in Computer Science, vol 11438*, pages 389–393, 2019a.
- Alípio M. Jorge, Ricardo Campos, Adam Jatowt, and Sérgio Nunes. Special issue on narrative extraction from texts (text2story): Preface. In *Information Processing Management an International Journal. Elsevier, Vol 56(5)*, pages 1771–1774, 2019b. URL <https://www.sciencedirect.com/science/article/abs/pii/S0306457319304455>.
- Vani Kanjirangat, Simone Mellace, and Alessandro Antonucci. Temporal embeddings and transformer models for narrative text understanding. In *Proceedings of Text2Story - 3rd Workshop on Narrative Extraction From Texts, co-located with the 42nd European Conference on Information Retrieval, Text2Story@ECIR 2020, Lisbon, Portugal, April 14th*, pages 71–77, 2020.
- O-Joun Lee and Jin-Taek Kim. Measuring narrative fluency by analyzing dynamic interaction networks in textual narratives. In *Proceedings of Text2Story - 3rd Workshop on Narrative Extraction From Texts, co-located with the 42nd European Conference on Information Retrieval, Text2Story@ECIR 2020, Lisbon, Portugal, April 14th*, pages 15–22, 2020.
- Mathis Linger and Mhamed Hajaiej. Batch clustering for multilingual news streaming. In *Proceedings of Text2Story - 3rd Workshop on Narrative Extraction From Texts, co-located with the 42nd European Conference on Information Retrieval, Text2Story@ECIR 2020, Lisbon, Portugal, April 14th*, pages 55–61, 2020.
- Chenyang Lyu, Tianbo Ji, and Yvette Graham. Incorporating context and knowledge for better sentiment analysis of narrative text. In *Proceedings of Text2Story - 3rd Workshop on Narrative Extraction From Texts, co-located with the 42nd European Conference on Information Retrieval, Text2Story@ECIR 2020, Lisbon, Portugal, April 14th*, pages 39–45, 2020.
- Martino Mensio, Harith Alani, and Alistair Willis. Towards a cross-article narrative comparison of news. In *Proceedings of Text2Story - 3rd Workshop on Narrative Extraction From Texts, co-located with the 42nd European Conference on Information Retrieval, Text2Story@ECIR 2020, Lisbon, Portugal, April 14th*, pages 89–94, 2020.
- Jakub Piskorski, Vanni Zavarella, Martin Atkinson, and Marco Verile. Timelines: Entity-centric event extraction from online news. In *Proceedings of Text2Story - 3rd Workshop on Narrative Extraction From Texts, co-located with the 42nd European Conference on Information Retrieval, Text2Story@ECIR 2020, Lisbon, Portugal, April 14th*, pages 105–114, 2020.
- Fabian Suchanek. The need to move beyond triples. In *Proceedings of Text2Story - 3rd Workshop on Narrative Extraction From Texts, co-located with the 42nd European Conference on Information Retrieval, Text2Story@ECIR 2020, Lisbon, Portugal, April 14th*, pages 95–104, 2020.