Event Report

Report on the 6th International Workshop on Narrative Extraction from Texts (Text2Story 2023) at ECIR 2023

Ricardo Campos  
INESC TEC & IPT  
Portugal  
ricardo.campos@ipt.pt

Alípio M. Jorge  
INESC TEC & FCUP  
Portugal  
amjorge@fc.up.pt

Adam Jatowt  
Univ. of Innsbruck  
Austria  
jatowt@acm.org

Sumit Bhatia  
Media and Data Science Research Lab, Adobe  
India  
Sumit.Bhatia@adobe.com

Marina Litvak  
Shamoon College of Engineering  
Israel  
litvak.marina@gmail.com

João Paulo Cordeiro  
INESC TEC & Univ. of Beira Interior  
Portugal

Conceição Rocha  
INESC TEC  
Portugal

Hugo Sousa  
INESC TEC  
Portugal

Behrooz Mansouri  
University of Southern Maine  
USA

Abstract

The Sixth International Workshop on Narrative Extraction from Texts (Text2Story’23) was held on April 2nd, 2023, in conjunction with the 45th European Conference on Information Retrieval (ECIR 2023) in Dublin, Ireland. Continuing the tradition of past years, the workshop was held as a hybrid event. Online participation was allowed using the Zoom platform. During the course of the day, more than 50 attendees had the opportunity to follow up and discuss the recent advances in topics related to representation, extraction, and generation of narratives. The workshop program included two invited keynotes and nineteen paper presentations. The proceedings of the workshop are available online.

Date: 2 April 2023.
Website: https://text2story23.inesctec.pt/.

1https://ceur-ws.org/Vol-3370/
1 Introduction

The Text2Story workshop series was established with the primary objective of bringing together researchers from diverse but related fields such as Information Retrieval, Natural Language Processing, Computational Linguistics, Artificial Intelligence, Human-Computer Interaction, and Data Visualization to share the recent advances in their respective fields towards narrative understanding. Building upon the success of the past editions [Jorge et al., 2018, 2019a; Campos et al., 2020, 2021, 2022] and on the Text2Story Special Issue at IPM Journal [Jorge et al., 2019b], this year, we organized the sixth edition of the Text2Story workshop, held as a hybrid event under the umbrella of ECIR 2023. Figure 1 captures the organizers and a few of the 50 participants of the workshop.

Figure 1. Organizers and participants at the Text2Story 2023 Workshop coming together for a group picture.

This report provides an overview of the main activities that took place during the workshop. Despite the hybrid format of this year’s edition, with both physical and virtual attendance options, the attendees’ enthusiastic engagement ensured that it was just as intellectually stimulating as previous editions. A total of 32 valid submissions were received, with 9 papers accepted in the research track. In addition, 10 papers were accepted in the demos, work-in-progress, and dissemination track, providing a platform for researchers to share ongoing work and disseminate previously published works from other venues. The final program comprised 19 papers, along with
two invited keynote talks and a vibrant and participated demo session. The presentation slides and
the videos of the talks at the workshop can be found at https://text2story23.inesctec.pt/.

2 The Program

The workshop program consisted of two invited keynote talks and nine research track papers. In
addition, ten papers were selected in the demo and work-in-progress categories along with
dissemination papers to allow for wider sharing of published work in other venues. We now
present a brief overview of the works presented at the workshop.

2.1 Keynotes

Dr. Georgiana Ifrim, from University College Dublin, Ireland, shared her work on the problem
of creating structured summaries of news in the form of timelines [Ifrim, 2023]. Timelines are highly
useful for news consumers as they present the most important topics and events in chronological
order. Generating such timelines automatically from large-scale news corpora is a challenging
task. Dr. Ifrim provided an overview of the major challenges followed by a survey of high-level
approaches for timeline summarization. She then presented her work to address the problem and
concluded with various qualitative examples of the structured timeline summaries thus generated.

The second keynote by Valentina Bartalesi, a researcher at the CNR-ISTI, Italy, presented
the idea of Semantic Story Maps – online interactive maps enriched with text, pictures, videos,
and other multimedia information, that convey a story over a territory [Bartalesi, 2023]. She
presented their work on a semi-automatic workflow that utilizes Semantic Web technologies and
ontological information to produce semantic story maps from textual documents. Finally, var-
ious demonstrations of story maps covering diverse topics such as medieval journeys, biological
processes, and stories of historical legends were presented.

2.2 Full Papers Research Track

The papers in the full paper research track of the workshop spanned a variety of topics ranging
from fundamental advances in narrative models to applications in news, clinical domain, and
video recommendations. Costa and Nunes [2023] explored the idea of using visual representations
of narratives to aid better data understanding. They presented NewsLine, an open-source web-
based prototype, that produces narrative visualizations of news content and allows the end-users
to play with different elements of the narrative (events, actors, etc.) in various ways. The idea
of visual representation of narratives was also presented by Silvano et al. [2023]. They described
a comprehensive annotation scheme and showed with qualitative examples how it can be used to
tag and link key events in news items thereby helping the readers grasp how the events relate
to the underlying story. Gervás [2023] argued that events in a complex narrative often include
events beyond simple facts and often represent the worldview of different story characters. Thus,
events in a complex narrative may be reported in different modes. He presented a constituency
parsing-based approach to identify spans in the narrative text corresponding to different reporting
modes. Saiz and Altuna [2023] presented their approach for temporal relation extraction in the
clinical domain.
Liu et al. [2023] presented a novel task of classifying movies based on their story types. They developed a manually annotated dataset of 45 movies labeled based on their story type and presented a machine-learning solution to classify the movies into different story types. González-García et al. [2023] applied the concepts of semantic storytelling in the context of automated speech assistants. They augment the answers provided by the automated assistant by recommending additional information to be shared with end-user to help them better understand the answers and related information.

A number of papers in the track dealt with the important problem of biased and harmful narratives. Tavakoli et al. [2023] presented their analysis of the readability of true and false content and reported that incorrect claims on the web are often harder to read than true claims. Onyepumuka et al. [2023] presented a case study of YouTube’s video recommendation algorithm and observed specific biases in YouTube’s recommendation algorithm where the recommended videos slowly drift from the topic of the seed videos. Vanetik et al. [2023] studied the problem of detecting negative campaigns in elections – those that focus on negative aspects of the competing candidate. They report that, to some extent, the lack of training data for detecting negative campaigns in low-resourced languages, such as Hebrew, can be compensated for by utilizing datasets for detecting offensive language in the same or other languages.

2.3 Ongoing Research Efforts and Resource Papers

The workshop also included special tracks for work-in-progress, demos, resources, and position papers to offer a platform for authors to share initial results of their research efforts and gather useful feedback from the community, demonstrate their research prototypes, introduce datasets and discuss a research vision or a reasoned opinion on the workshop topics. The following ten papers were selected in this track.

2.3.1 Work-in-Progress

Ruggeri et al. [2023] provides a comprehensive overview of the challenges of subjectivity detection and presents a case study to prove how a prescriptive annotation approach can improve language-agnostic subjectivity detection results. Kanjirangat and Antonucci [2023] describes a work in progress for the unsupervised labeling of edges in narrative knowledge graphs. They detect edges between two entities (characters in a novel) based on the verbs in the sentences. To label the detected edges they utilize a combination of word embedding based clustering and synsets from WordNet. Sami et al. [2023a] attempted to analyze a fundamental problem in storytelling, that is to help users navigate arguments presented in a news article. In this regard, they presented a theory of mapping 5Ws (Who, Where, What, When and Why) and Aristotle’s Rhetoric into the format of Joseph Campbell’s The Hero’s Journey as a structural story template to assist in automatic understanding of the structure of news and evaluated the approach via cognitive reading and writing user experiment tasks.
2.3.2 Position Papers

Palma [2023] aims to attract attention towards modeling the story’s interestingness and to initiate a discussion about the re-framing of computational narratology in its needs. In this regard, the author suggests modeling stories as L-Systems and Magic Squares.

2.3.3 Resource Papers

Alrashid and Gaizauskas [2023] presents a new dataset, the ScANT corpus, containing the annotation of 14 chapters, from 6 sources (two children’s books and four adult novels), using the SceneML framework, an annotation language developed by the authors for annotating scenes in narrative text. The annotation was made by two native English-speaking PhD students in English Literature and Computational Linguistics. About 2,800 sentences were annotated and analyzed according to the classes of the guidelines. Piskorski et al. [2023] presents a new corpus consisting of circa 4.5K news snippets labeled with fine-grained infectious disease-related event types, which can serve as a resource for training and building ML-based models for event detection/classification in epidemics-related online news. The manual annotation was performed by 7 experienced annotators following a consistent approach using a codebook.

2.3.4 Demo Papers

Ezeani et al. [2023] presents the application of NLP tools on annotating and mining document collections primarily in the area of digital humanities to obtain a humanistic point of view of imprecise geographical locations and their relations with respect to time and human sentiment. Experiments were made using the English Lake District corpus. Veale [2023] describes a system called Excelsior! that generates engaging data-driven comic strips that integrate two sides of a particular argument into a single visual narrative. The objective is to expose people to both sides of the argument and prevent current challenges like echo chambers and factionalism on social media platforms. The paper outlines the use of comic strips as a narrative medium to balance two points of view. Sami et al. [2023b] describes a demo system, named Terms Board (TB), that can be used to help writers in story planning. The system works over a fixed collection of documents (e.g. a news collection) and the user can insert a query to obtain results. The results are presented in a graphical interface, organized in 6 different panels (who, where, what-topics, what-actions, why-positives, and why-negatives). To evaluate the system for story writing, an online experiment with 32 participants was organized. Finally, Bartalesi et al. [2023] presents the Story Map Building and Visualizing Tool, a user-friendly web interface that allows a user to input the subject of a story which is then populated with suggestions for entities in Wikipedia and Wikidata. The demonstration laid out in the paper shows the following three scenarios:

1. allowing the user to pick relevant aspects of the story from the recommended entities and other suggestions;
2. search-based functionality to retrieve relevant stories; and
3. finally a geo-spatial map-based interface for portraying relevant storylines of an entity or an event.
3 Key Contributors

The Program Chairs were Ricardo Campos, Alípio Jorge, Adam Jatowt, Sumit Bhatia, and Marina Litvak. The proceedings were setup by our Proceedings Chairs João Paulo Cordeiro and Conceição Rocha. Hugo Sousa and Behrooz Mansouri were the web and dissemination chairs. Jochen Leidner (Coburg University) and Purificação Silvano (CLUP & University of Porto) also chaired technical sessions at the workshop.

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’23 workshop. Our thanks go to:

- Álvaro Figueira (INESC TEC & University of Porto)
- Andreas Spitz (University of Konstanz)
- Antoine Doucet (Université de La Rochelle)
- António Horta Branco (University of Lisbon)
- Anubhav Jangra (IIT Patna)
- Arian Pasquali (Faktion AI)
- Bart Gajderowicz (University of Toronto)
- Begona Altuna (Universidad del País Vasco)
- Behrooz Mansouri (Rochester Institute of Technology)
- Brenda Santana (Federal University of Rio Grande do Sul)
- Bruno Martins (IST & INESC-ID, University of Lisbon)
- Daniel Loureiro (Cardiff University)
- Dennis Aumiller (Heidelberg University)
- Dhruv Gupta (Norwegian University of Science and Technology)
- Dyaa Albakour (Signal UK)
- Evelin Amorim (INESC TEC)
- Henrique Lopes Cardoso (LIACC & University of Porto)
- Hugo Sousa (INESC TEC & University of Porto)
- Ismail Altingovde (Middle East Technical University)
- Irina Rabaev (Shamoon College of Engineering)
- Jiexin Wang (South China University of Technology, China)
- João Paulo Cordeiro (INESC TEC & University of Beira Interior)
- Kiran Bandeli (Walmart Inc.)
- Liana Ermakova (HCTI, Université de Bretagne Occidentale)
- Luca Cagliero (Politecnico di Torino)
- Ludovic Moncla (INSA Lyon)
- Luís Filipe Cunha (INESC TEC & University of Minho)
- Marc Finlayson (Florida International University)
- Marc Spaniol (Université de Caen Normandie)
- Moreno La Quatra (Politecnico di Torino)
- Natalia Vanetik (Shamoon College of Engineering)
- Nuno Guimarães (INESC TEC & University of Porto)
- Pablo Gamallo (University of Santiago de Compostela)
- Pablo Gervás (Universidad Complutense de Madrid)
4 Recognition Awards

The workshop ended with the announcement of the recognition awards. The Text2Story 2023 Best Paper Award went to the paper entitled *Multi-label Infectious Disease News Event Corpus* by Piskorski et al. [2023]. In addition to this, the workshop chairs have recognized the following researchers as recipients of the 2023 reviewer award for their insightful and valuable reviews.

- Andreas Spitz
- Bruno Martins
- Dennis Aumiller
- Dhruv Gupta
- Marc Finlayson
- Pablo Gervas
- Purificação Silvano
- Sérgio Nunes
- Sriharsh Bhyravajjula
- Yang Zhang

5 Concluding Thoughts

The Text2Story workshop series has reached its sixth edition. Our initial goal was to unite participants with diverse backgrounds and expertise in order to collaboratively establish the direction for the emerging field of *extracting narratives from texts*. We are delighted by the overwhelmingly positive response from the research community, and the increasing number of participants in the workshops highlights the importance of the topic.
Acknowledgements

Our gratitude goes to the ECIR 2023 organizers for their assistance in facilitating the workshop, as well as our exceptional keynote speakers, program committee members, paper authors, and participants whose contributions led to the tremendous success of the workshop.

References


Ricardo Campos, Alípio Jorge, Adam Jatowt, Sumit Bhatia, and Marina Litvak. The 5th international workshop on narrative extraction from texts: Text2story 2022. In Matthias


