



The 3rd International Workshop on Narrative Extraction from Texts: Text2Story 2020

Ricardo Campos^{1,2} , Alípio Jorge^{1,3} , Adam Jatowt⁴ ,
and Sumit Bhatia⁵ 

¹ INESC TEC, Porto, Portugal

² Ci2 - Smart Cities Research Center - Polytechnic Institute of Tomar,
Tomar, Portugal
`ricardo.campos@ipt.pt`

³ University of Porto, Porto, Portugal
`amjorge@fc.up.pt`

⁴ Kyoto University, Kyoto, Japan
`adam@dl.kuis.kyoto-u.ac.jp`

⁵ IBM Research AI, New Delhi, India
`sumitbhatia@in.ibm.com`

Abstract. The Third International Workshop on Narrative Extraction from Texts (Text2Story'20) [text2story20.inesctec.pt] held in conjunction with the 42nd European Conference on Information Retrieval (ECIR 2020) gives researchers of IR, NLP and other fields, the opportunity to share their recent advances in extraction and formal representation of narratives. This workshop also presents a forum to consolidate the multi-disciplinary efforts and foster discussions around the narrative extraction task, a hot topic in recent years.

Keywords: Information extraction · Narrative extraction

1 Background and Motivation to ECIR

Searching for relevant information is a permanent need for those who want to stay informed about a given event, news or story. While having access to information is now easier than ever with the proliferation of devices and different means of accessing data, keeping up-to-date with all the developments and various aspects of the topic being followed is a difficult task. In many situations, it is hard for readers to connect the dots of a given story [21]. This is due not only to the widespread presence of the media outlets in the digital space [14], but also to the increasing participation of citizens, who produce and promote an unprecedented number of comments and discussions on social media (some of them lasting over days, weeks or months). Automatic narrative extraction from texts offers a compelling approach to this problem by automatically identifying the sub-set of

interconnected raw documents, extracting the critical narrative/story elements, and representing them in a more adequate manner that conveys the key points of the story in an easy to understand format to the readers. This could be done through text summarization [13], timelines [15, 19], word clouds [3, 20], visual textual analytics [7, 18] or in an intermediate structured formalism (e.g., wiki-like page structures [1]) that can feed further steps (e.g., gamification [6, 16] or story generation [12] (such as automatically generating finance [5, 11] and sport reports [22])).

Although information extraction and natural language processing have made significant progress towards automatic interpretation of texts, the problem of fully identifying and relating the different elements of a narrative in a document (set) still presents significant unsolved challenges [17].

The purpose of the Text2Story workshop series is to shorten the distance between IR and people working on automatic narrative extraction and construction from texts, a vibrant line of research that has been conducted over the last few years by many research groups.

2 Past-Related Activities

The Text2Story workshop series had its first edition at ECIR'18 [8], followed by a second edition on ECIR'19 [9]. In these two first editions, we had an approximate number of 70 participants, 16 research papers presented, plus demo and poster sessions, and vibrant talks from our four invited keynotes: Udo Kruschwitz (University of Essex), Eric Gaussier (University Grenoble Alps), Iryna Gurevych (Technische Universität Darmstadt), and Miguel Martinez-Alvarez (Signal AI). In addition to this, we also edited the Text2Story Special Issue on IPM Journal [10] which had more than 30 submissions and 8 papers accepted, demonstrating the growing activity of this specific research area. The organizers of the workshop have also been actively involved in this research area with the proposal and the contribution of new methods and solutions, most notably the [YAKE!](#) keyword extraction algorithm [2–4] - best short paper of ECIR'18 - and the [Tell me Stories](#) temporal summarization tool [15] - best demo presentation at ECIR'19.

In the third edition of this workshop series, we aim to raise awareness to the problem of creating text-to-narrative-structures and its related tasks. We focus on researchers and practitioners working on identifying, extracting and producing narrative stories, but also on people from industry, particularly journalists and stakeholders working in traditional and social media.

3 Topic Outline

The call for papers aimed to cover original research at the intersection of IR and NLP on all aspects of storyline identification and generation from texts including but not limited to narrative and content generation, formal representation, and visualization of narratives. The topics of the workshop are in line with the

previous editions of the Text2Story workshop series. In particular, we featured the following topics:

- Event Identification
- Narrative Representation Language
- Information Retrieval Models based on Story Evolution
- Narrative-focused Search in Text Collections
- Temporal Information Retrieval and Narrative Extraction
- Sentiment and Opinion Detection
- Argumentation Mining
- Narrative Summarization
- Multi-modal Summarization
- Storyline Visualization
- Temporal Aspects of Storylines
- Story Evolution and Shift Detection
- Causal Relation Extraction and Arrangement
- Evaluation Methodologies for Narrative Extraction
- Big data applied to Narrative Extraction
- Resources and Dataset showcase
- Personalization and Recommendation
- User Profiling and User Behavior Modeling
- Credibility
- Models for detection and removal of bias in generated stories
- Ethical and fair narrative generation
- Fact Checking
- Bots Influence
- Bias in Text Documents
- Automatic Timeline Generation

4 Program Committee and Support Chairs

All papers were refereed through a double-blind peer-review process by at least three reviewers and are planned to be published by [CEUR](#). The program committee members consist of researchers from industry and academia. The following members formed the program committee of the Text2Story’20 workshop:

- Álvaro Figueira (INESC TEC & University of Porto)
- Andreas Spitz (École polytechnique fédérale de Lausanne)
- António Horta Branco (University of Lisbon)
- Arian Pasquali (Signal AI)
- 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 AI)
- 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 (Universidade de 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)

Proceedings Chair

- João Paulo Cordeiro (INESC TEC; Universidade da Beira do Interior, Covilhã, Portugal)
- Conceição Rocha (INESC TEC)

Web and Dissemination Chair

- Arian Pasquali (Signal AI)
- Behrooz Mansouri (Rochester Institute of Technology)

Acknowledgements. The first two authors of this paper are 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

1. Alonso, O., Kandylas, V., Tremblay, S.-E.: How it happened: discovering and archiving the evolution of a story using social signals. In: Proceedings of JCDL 2018, pp. 193–202. ACM (2018). <https://doi.org/10.1145/3197026.3197034>

2. Campos, R., Mangaravite, V., Pasquali, A., Jorge, A.M., Nunes, C., Jatowt, A.: A text feature based automatic keyword extraction method for single documents. In: Pasi, G., Piwowarski, B., Azzopardi, L., Hanbury, A. (eds.) ECIR 2018. LNCS, vol. 10772, pp. 684–691. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-76941-7_63
3. Campos, R., Mangaravite, V., Pasquali, A., Jorge, A.M., Nunes, C., Jatowt, A.: YAKE! collection-independent automatic keyword extractor. In: Pasi, G., Piwowarski, B., Azzopardi, L., Hanbury, A. (eds.) ECIR 2018. LNCS, vol. 10772, pp. 806–810. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-76941-7_80
4. Campos, R., Mangaravite, V., Pasquali, A., Jorge, A., Nunes, C., Jatowt, A.: YAKE! keyword extraction from single documents using multiple local features. *Inf. Sci.* **509**, 257–289 (2020)
5. El-Haj, M., Rayson, P., Moore, A.: The first financial narrative processing workshop (FNP 2018). In: Proceedings of the LREC 2018 Workshop (2018)
6. Grobelny, J., Smierzchalska, J., Krzysztof, K.: Narrative gamification as a method of increasing sales performance: a field experimental study. *Int. J. Acad. Res. Bus. Soc. Sci.* **8**(3), 430–447 (2018). <https://doi.org/10.6007/IJARBSS/v8-i3/3940>
7. Hoque, E., Carenini, G.: MultiConVis: a visual text analytics system for exploring a collection of online conversations. In: Proceedings of IUI 2016, pp. 96–107. ACM (2016). <https://doi.org/10.1145/2856767.2856782>
8. Jorge, A., Campos, R., Jatowt, A., Nunes, S.: First international workshop on narrative extraction from texts (Text2Story 2018). In: Pasi, G., Piwowarski, B., Azzopardi, L., Hanbury, A. (eds.) ECIR 2018. LNCS, vol. 10772, pp. 833–834. Springer, Heidelberg (2018). <https://doi.org/10.1007/978-3-319-76941-7>
9. Jorge, A., Campos, R., Jatowt, A., Bhatia, S.: Second international workshop on narrative extraction from texts (Text2Story 2019). In: Azzopardi, L., Stein, B., Fuhr, N., Mayr, P., Hauff, C., Hiemstra, D. (eds.) ECIR 2019. LNCS, vol. 11438, pp. 389–393. Springer, Heidelberg (2019). https://doi.org/10.1007/978-3-030-15719-7_54
10. Jorge, A., Campos, R., Jatowt, A., Nunes, S.: Special issue on narrative extraction from texts (Text2Story): preface. *Inf. Process. Manage. Int. J.* **56**(5), 1771–1774 (2019). <https://doi.org/10.1016/j.ipm.2019.05.004>
11. Lewis, C., Young, S.: Fad or future? Automated analysis of financial text and its implications for corporate reporting. *Account. Bus. Res.* **49**(5), 587–615 (2019). <https://doi.org/10.1080/00014788.2019.1611730>
12. Li, B., Lee-Urban, S., Johnston, G., Riedl, M.: Story generation with crowdsourced plot graphs. In: Proceedings of the AAAI 2013, pp. 598–604. AAAI Press (2013). <https://doi.org/10.5555/2891460.2891543>
13. Liu, S., et al.: TIARA: interactive, topic-based visual text summarization and analysis. *ACM Trans. Intell. Syst. Technol.* **3**(2), 1–28 (2012). Article no. 25. <https://doi.org/10.1145/2089094.2089101>
14. Martínez-Alvarez, M., et al.: First international workshop on recent trends in news information retrieval (NewsIR 2016). In: Ferro, N., et al. (eds.) ECIR 2016. LNCS, vol. 9626, pp. 878–882. Springer, Heidelberg (2016). https://doi.org/10.1007/978-3-319-30671-1_85
15. Pasquali, A., Mangaravite, V., Campos, R., Jorge, A., Jatowt, A.: Interactive system for automatically generating temporal narratives. In: Azzopardi, L., Stein, B., Fuhr, N., Mayr, P., Hauff, C., Hiemstra, D. (eds.) ECIR 2019. LNCS, vol. 11438, pp. 251–255. Springer, Heidelberg (2019). https://doi.org/10.1007/978-3-030-15719-7_34

16. Pujolà, J-T., Argüello, A.: Stories or scenarios: implementing narratives in gamified language teaching. In: Arnedo-Moreno, J., González, C.S., Mora, A. (eds.) *GamiLearn 2019* (2019)
17. Riedl, O.: Computational narrative intelligence: a human-centered goal for artificial intelligence. [arxiv.org](https://arxiv.org/abs/2016) (2016)
18. Seifert, C., Sabol, V., Kienreich, W., Lex, E., Granitzer, M.: Visual analysis and knowledge discovery for text. In: Gkoulalas-Divanis, A., Labbi, A. (eds.) *Large-Scale Data Analytics*, pp. 189–218. Springer, New York (2014). https://doi.org/10.1007/978-1-4614-9242-9_7
19. Tran, G., Alrifai, M., Herder, E.: Timeline summarization from relevant headlines. In: Hanbury, A., Kazai, G., Rauber, A., Fuhr, N. (eds.) *ECIR 2015. LNCS*, vol. 9022, pp. 245–256. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-16354-3_26
20. Viégas, F., Wattenberg, M.: Tag clouds and the case for vernacular visualization. *Interactions* **15**(4), 49–52 (2008). <https://doi.org/10.1145/1374489.1374501>
21. Vossen, P., Caselli, T., Kontzopoulou, Y.: Storylines for structuring massive streams of news. In: *Proceedings of the First Workshop on Computing News Storylines (CNewsStory 2015@ACL-IJCNLP 2015)*, pp. 40–49. ACL (2015). <https://doi.org/10.18653/v1/W15-4507>
22. Zhang, J., Yao, J.-G., Wan, X.: Towards constructing sports news from live text commentary. In: *Proceedings of ACL 2016*, pp. 1361–1371. ACL (2016). <https://doi.org/10.18653/v1/P16-1129>