



Interactive System for Automatically Generating Temporal Narratives

Arian Pasquali, Vítor Mangaravite, Ricardo Campos
Alípio Jorge and Adam Jatowt

Universidade do Porto, Instituto Politécnico de Tomar e Universidade de Kyoto



Awarded 1st Prize at Portuguese Web Archive 2018 contest

<http://archive.tellmestories.pt>

ECIR Demo

<http://demo.tellmestories.pt>



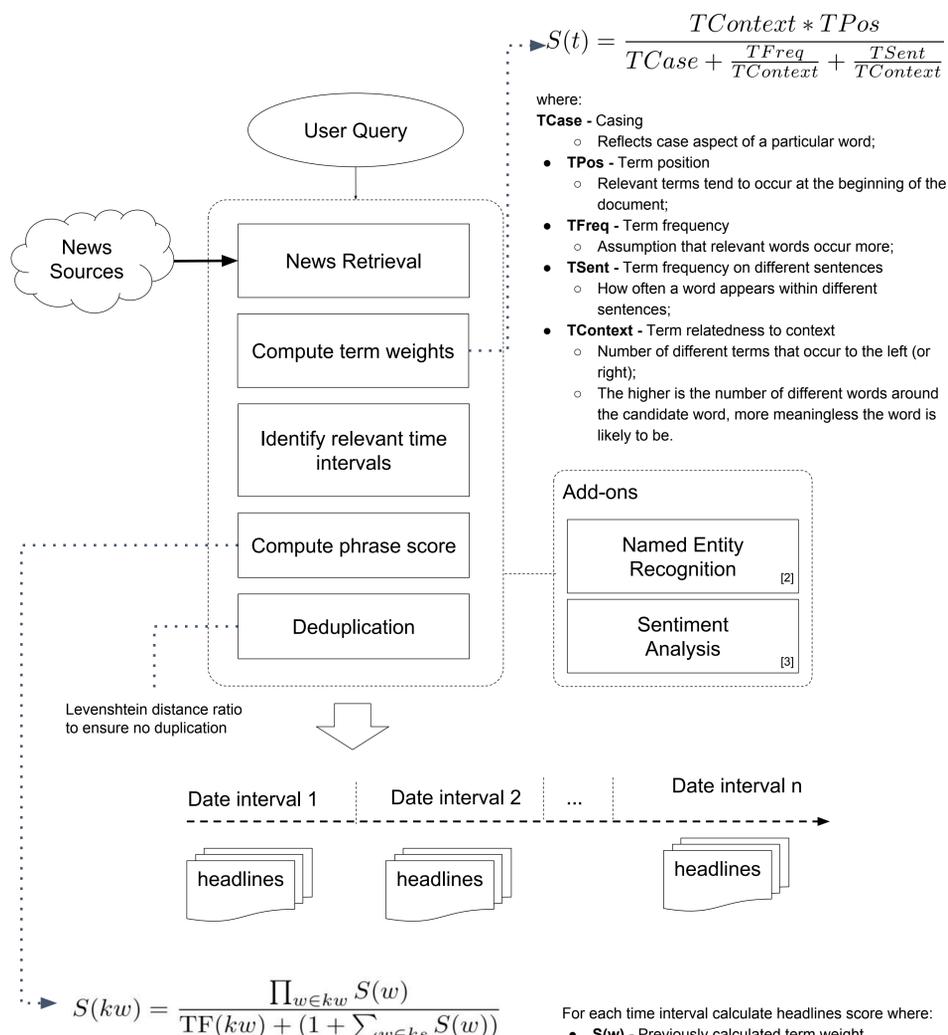
Conta-me Histórias (Tell me stories) is an online tool that allows users to automatically generate an interactive temporal summarization based on search queries.

Our approach is based on a Keyphrase Extraction algorithm [1] to improve ranking precision and a peak detection method to select relevant headlines over long periods of time. It offers a friendly user interface that enables users to study and revisit topics in the past thus providing a different perspective on historical narratives.

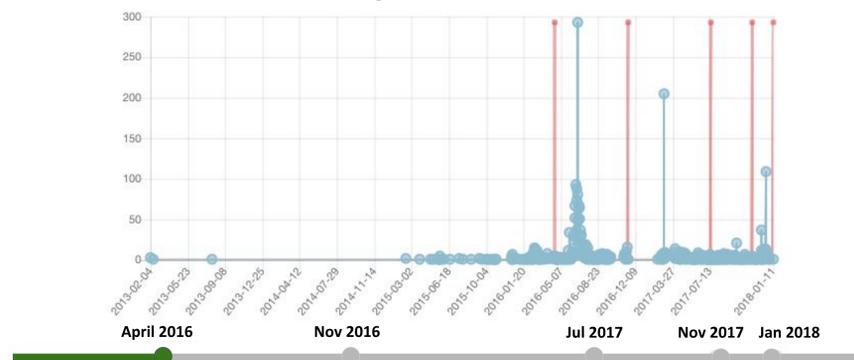
The architecture described here is an extendable framework to identify relevant headlines and important periods that can be applied to any dataset of news with timestamped data. We made available at <http://demo.tellmestories.pt> an instance that allows users to explore different datasets like the Portuguese Web Archive, Bing News and the The Signal Media One-Million News Articles Dataset [2].

Figure 2 illustrates time interval selection for the query “Brexit” considering the last 10 years of news in the Portuguese Web Archive. The system identified five important time intervals between 2013 and 2018. The red lines represent interval boundaries, while the blue area highlights the number of news aggregated by date.

Architecture



Timeline summarization for **Brexit** during the last 10 years in the Portuguese media



Headlines between **February 4th 2013** to **April 19th 2016**

- UK secretly studies a way out European Union
- Immigration creates impasse between UK and Brussels.
- Brussels gives way to Cameron to avoid Brexit.
- Can the UK survive Brexit?
- Rajoy confident in agreement between EU and UK.
- British businessmen next to Cameron against Brexit.
- Brexit or not Brexit.
- Boris Johnson challenges Cameron and will campaign for brexit.
- Cameron ridicules Boris Johnson's decision to support Brexit
- Juncker receives António Costa next week in Brussels before summit.
- Cameron challenges Brexit's defenders to take risks of leaving.
- Brexit could cause a decade of uncertainty in Britain

Figure 2. Result sample of an interactive timeline. Headlines were translated from Portuguese to English

Conclusions and next steps

In the era of post-truth and fake news, web archive and media monitoring initiatives are important contributions to preserve history. We believe that making this demo publicly available and accessible for everyone is an important contribution to foster not only related research, but also the user's search experience when looking for past events and summarizing complex information.

This is a preliminary study in this field and the next steps involve extended evaluation to assess summarization quality and experiments with different timeline segmentation methods. In the future we hope to be able to generate temporal summarization results taking into consideration aspects like bias, hyperpartisan argumentation and trust. Figure 3 illustrates our preliminary results in bias classification using annotated corpus from SemEval 2019 [3].

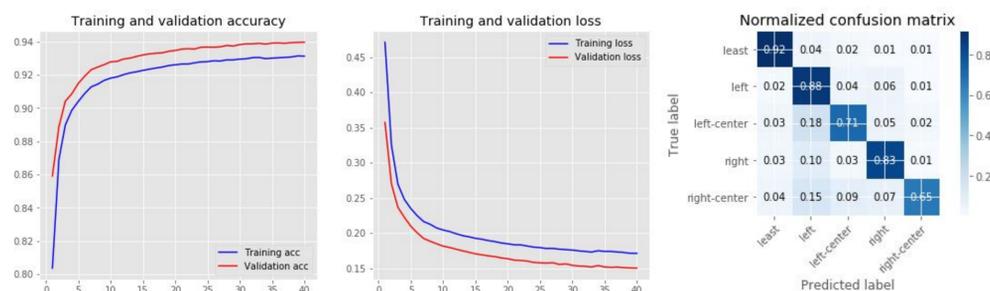


Figure 3. Bias classifier performance results using 600.000 annotated corpus. Classes [none, left, center-left, center-right, right]

Contributions



- Code is open source and extendable. Anyone can extend and use it with different datasets.

<http://opensource.tellmestories.pt>

- Friendly user interface for temporal summarization for users of the Portuguese Web Archive

<http://archive.tellmestories.pt>

- Demo using different datasets and methods

<http://demo.tellmestories.pt>

References

- Campos, R. and Mangaravite, V. and Pasquali, A. and Jorge, A.M. and Nunes, C. and Jatowt, A.: A Text Feature Based Automatic Keyword Extraction Method for Single Documents. ECIR, 684–691 (2018). Best short paper ECIR 2018. (<http://yake.inesctec.pt>)
- D. Comey, D. Albakour, M. Martinez, S. Moussa. What do a Million News Articles Look like? ECIR 2016 (The Signal Media One-Million News Articles Dataset)
- SemEval 2019 Task Hyperpartisan News Detection

Acknowledgements

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.

