



Conta-me Histórias: Automatically Generating Interactive Temporal Summarization in Portuguese

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

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



Awarded 1^o Prize at Arquivo.pt 2018 contest

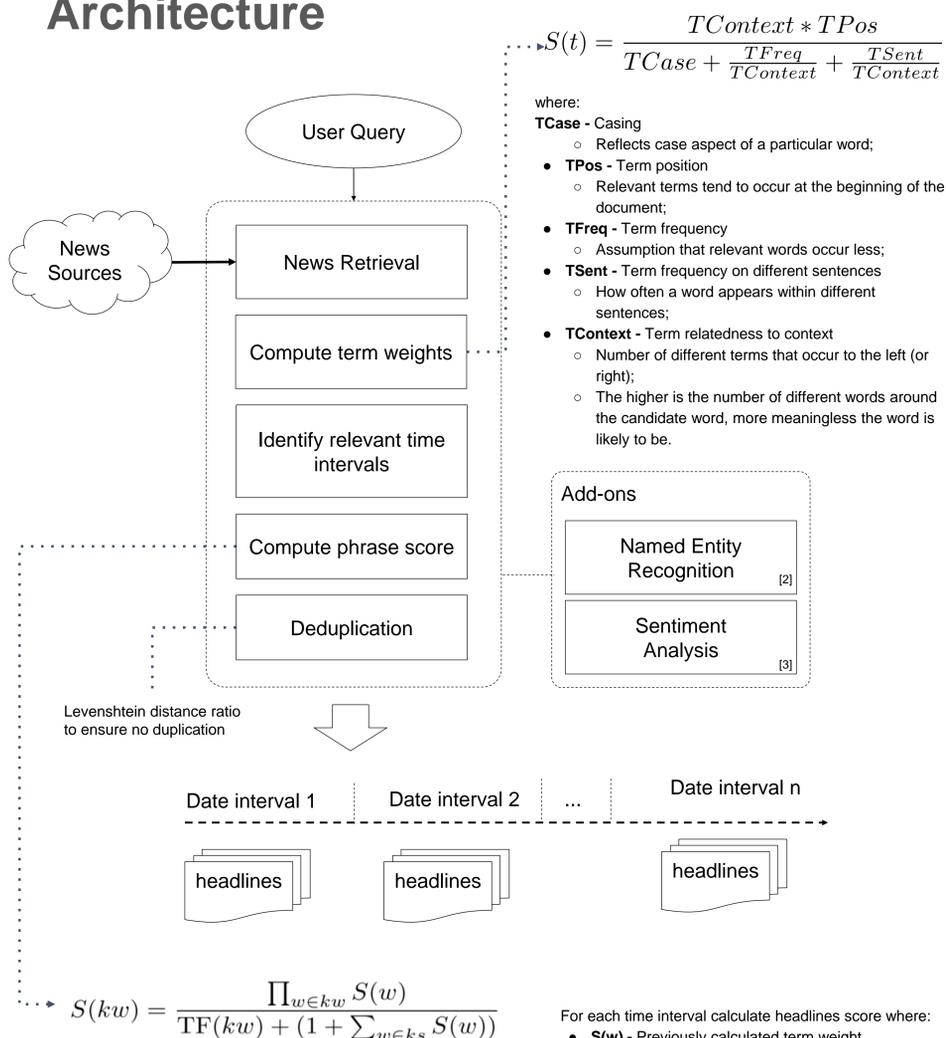
<http://contamehistorias.pt>

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

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.

Architecture



Example

Figure 2 illustrates a timeline using query “Dilma Rousseff” considering the last 10 years of news in the Portuguese newspapers.



Figure 2. Result sample of an interactive timeline

Conclusions

In the era of post-truth and fake news, web archive initiatives are important contributions to preserve history. However, we should think about better ways to explore its content. In this context, our demo may be considered an additional solution that allows users to better explore this kind of repository.

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.

Despite using a web archive as data source, this tool can be adapted to support different kinds of data sources and languages.

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>)
- Rocha, C. and Jorge, A. and Sionara, R. and Brito, P. and Pimenta, C. and Rezende, S. PAMPO: using pattern matching and pos-tagging for effective Named Entities recognition in Portuguese, (2016)
- Silva, M. J. and Carvalho, P. and Costa, C and Sarmento L.: Automatic Expansion of a Social Judgment Lexicon for Sentiment Analysis. Technical Report TR 10-08. University of Lisbon, Faculty of Sciences, LASIGE, (2010)

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.

Contributions



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

2. Framework supports Portuguese and other languages.

3. Friendly user interface for temporal summarization for users of the Portuguese web archive (<http://arquivo.pt>)

<http://bit.ly/ContameCode>

