

Miguel Won (Instituto de Engenharia de Sistemas e Computadores - Investigação e Desenvolvimento (INESC-ID)), *The 'Arquivo de Opinião' newspaper archive*

Political commentary is a significant presence on news media. It's from the political commentary (through the so-called pundits) that many political events are selected as relevant, and consequently (re)interpreted in the process of the narrative construction of the "public realm". While the daily news reports events as they happen, the political commentators' role is to openly give his/her personal input about how such narrative should be told. This quality gives the political commentator a special and unique position in the news spread process, particularly as a main influencer of the public opinion. This idea is more clear when we often see news generated not by the political events but from the expressed opinions about such events, namely from highly rated political commentators. This privileged position gives these political commentators a special role in public debate, by deciding what and how should be discussed and interpreted. Furthermore, due to special visibility that is given by the news media as well the non-existence of direct contradictory, the political commentator position is a privileged place to spread political propaganda. This results in a close proximity between the political commentary world and the politicians' networks, often resulting in overlaps. This proximity is observed when political commenters are former, potential or in even in some cases active politicians.

"Arquivo de Opinião" is an online archive of such opinions, aimed to create a memory of the daily opinion published in the main Portuguese newspapers. An existing memory of this public debate sector is a contribution to the quality of the public debate. The possibility to remember older debates, what was written, who wrote about what or what were the main argumentative trends in a particular polarized issue can bring powerful scrutiny tools that help to improve the public debate. Due to the high volume of published articles, such archive should not only provide access to the articles but also offer user-friendly search tools, such as search engines or author and tags indexation.

"Arquivo de Opinião" has collected more than 80000 articles, written approximately by 3500 distinct authors, and published by the most relevant Portuguese newspapers in the period of 2008-2016. The archive offers a search engine where the user can search for words or expression, and filter the results by date, authorship or publication source. Every article was processed through an NLP pipeline that resulted in the indexation of more than 30000 relevant keywords and mentioned entities (geographic locations, person names and organizations). Additionally, every author has its own webpage that informs about the most relevant keywords used by the author, together with the most used mentioned entities. Furthermore, a search tool is available that allows the user to analyze trends of keywords, by time and media source, as well as relevant keywords used in the same context. "Arquivo de Opinião" collection is an innovative corpus, that offers the possibility for studies in several fields, from Political Science to Communication Studies.

*This research was supported by Fundação para a Ciência e Tecnologia (FCT), through the scholarship with reference SFRH/BPD/104176/2014, as well as through the INESC-ID multi-annual funding from the PIDDAC programme, which has the reference UID/CEC/50021/2013, and FEDER under the project 22153-01/SAICT/2016.*