**Augmented Democracy: A System for Digital Democracy\***

**Enzo Bustos da Silva**

**Thomas Palmeira Ferraz, André Seidel Oliveira**

**Anna Helena Reali Costa**

Escola Politécnica da Universidade de São Paulo

[enzobustos@usp.br](mailto:enzobustos@usp.br)

**Objectives**

With the increasing spread of the Internet, democratic participation has improved both in information processing as in communication and transactions. For example, citizens can obtain information about local politics and evaluate the evolution of their elected representatives; finally, acquire more and better information to guide their votes and political action. This work proposes Augmented Democracy, a system applied to digital democracy [1] that employs machine learning (ML) techniques and natural language processing (NLP) that increases the transparency of the democratic process. This was achieved through the automatic interpretation of documents generated in public bodies to translate these extensive and complex texts into something that any citizen more easily understands.

More specifically, these techniques were investigated in the minutes of the Diário da Assembleia da República Portuguesa1 (DAR), aiming at three contributions:

1. Process and structure the data produced by the legislative body;
2. Extract and process relevant information within political discourse - in particular, with automatic summarization, topic modeling, and sentiment analysis;
3. Provide forms of interpretation, visualization, and interaction that facilitate the understanding of the ordinary citizen.

**Materials and Methods**

To create this system, first, it was necessary to use web crawling and web scraping methods to collect the minute's data directly from the website of the Portuguese Parliament. This raw data passes through a pre-segmentation step to divide the entire text of the set of Minutes into a structured database, as shown in Figure 1. This process facilitates the processing and extraction of information from this data.

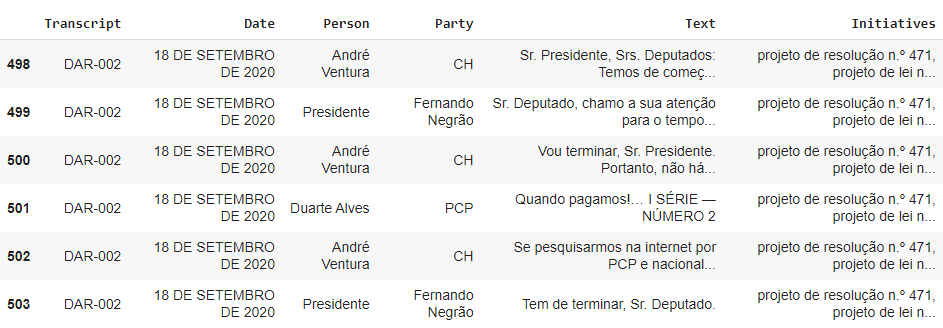


Figure 1: Data generated from DAR minutes.

In this data, it is worth mentioning the columns "Text" and "Initiatives". The first corresponds to each of the lines that took place in the plenary session, divided by speaker; the second deals with which agenda item corresponds to that statement (for example, a Project Resolution).

From this structured data, the user's interactions with the system follow the scheme in Figure 2. The user inputs pertinent information, such as the period of interest, a specific party or politician, or a particular subject of interest.

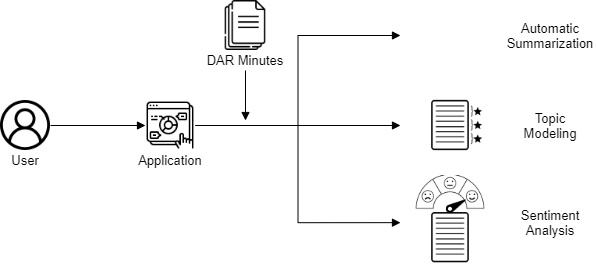


Figure 2: Scheme of the proposed Augmented Democracy system, with automatic summarization features, the definition of the topic addressed, and analysis of the feelings of the speeches.

In conjunction with the information provided by the user, sophisticated ML algorithms are applied, such as BERT [2] and its state-of-the-art variants, in order to perform three specialized functionalities:

1. Automatic Summarization [3]: aims to produce a summary from a set of input documents;
2. Topic Modeling [4]: method to identify sets of words to determine the topic covered in a given text;
3. Sentiment Mining [5]: a task that aims to determine the polarity of the feeling expressed in a text;

For this work, we used (1) to reduce the size of discussions on a particular agenda item; (2) as a search engine, guiding the user to discussions that are of interest to them and (3) to make sure that the arguments presented by a party are consistent with their vote. The work is still in progress and is expected to end in February 2022.

**Results**

As a result of this scientific initiation project, we have a unified system that encompasses several ML and NLP functionalities applied in the context of the Minutes of the Diário da Assembleia da República Portuguesa.

One of the partial results can be seen in Figure 3, in which the topic modeling algorithm finds the main words related to the same subject.

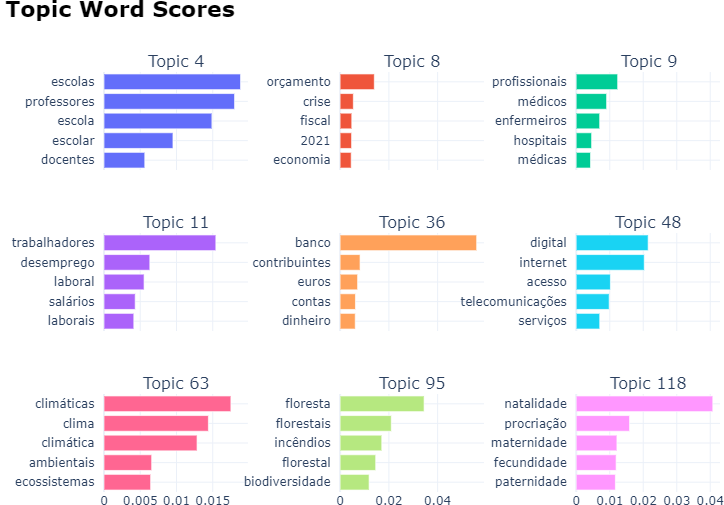


Figure 3: Example of the main topics identified by BERTopic in the minutes.

Another result, now from the automatic summarization, can be seen in Figure 4, an intervention text2 of 7903 characters is reduced to just one paragraph without considerable losses in the entire speech sense.

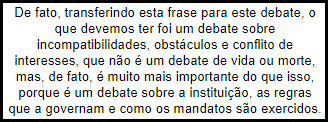


Figure 4: Example of an automatic summary generated by PEGASUS using abstractive techniques.

**Conclusions**

The Augmented Democracy system proposed in this summary aims to contribute mainly to the information axis of digital democracy. In addition to artificial intelligence techniques, we plan to create an application as future work, always aiming at a higher quality for the information received by the user.

In addition to contributing to other projects that want to go deeper into the other axes, this work also contributed three articles to the NLP area. [6] deals with techniques for approaching moderate debates, [7] deepens in a classification method after Topic Modeling, and [8] deals specifically with the homonymous tool to this summary for combating disinformation..

**References**

*[1] Breindl, Y. et al. "Can Web 2.0 applications save e-democracy? A study of how new internet applications may enhance citizen participation in the political process online". IJED (2008).*

*[2] Devlin, J. et al. "Bert: Pre-training of deep bidirectional transformers for language understanding". arXiv preprint (2018).*

*[3] Huang, D. et al. "What Have We Achieved on Text Summarization?". arXiv preprint (2020).*

*[4] Grootendorst, M. “BERTopic: Leveraging BERT and c-TF-IDF to create easily interpretable topics”. (2020)*

*[5] Abercrombie, G. et al. "‘Aye’ or ‘no’? Speech-level sentiment analysis of Hansard UK parliamentary debate transcripts." LREC (2018)*

*[6] Ferraz, T. et al. “DEBACER: a method for slicing moderated debates”. ENIAC (2021)*

*[7] Alcoforado, A. et al. "ZeroBERTo - Leveraging Zero-Shot Text Classification by Topic Modeling".*

*[8] Alcoforado, A. et al. “Augmented Democracy: Artificial Intelligence as an Ally in Fighting Disinformation”.*