**Democracia Aumentada: Um Sistema para Democracia Digital\***

**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)

**Objetivos**

Com a crescente disseminação da Internet, a participação democrática tem sido aprimorada tanto no processamento de informações, como na comunicação e transações. Por exemplo, o cidadão pode obter informações sobre a política local e avaliar a atuação de seus eleitos, além de nortear seus votos e sua atuação política. Este trabalho propõe a Democracia Aumentada, um sistema aplicado à democracia digital [1] que emprega técnicas de aprendizado de máquina (AM) e processamento de linguagem natural (PLN) visando aumentar a transparência do processo democrático. Isso foi feito através da interpretação automática dos documentos gerados em órgãos públicos, com a finalidade de traduzir esses textos, extensos e de difícil interpretação, em algo que seja mais facilmente compreendido por qualquer cidadão.

Mais especificamente, técnicas de AM e PLN foram investigadas nas atas do Diário da Assembleia da República Portuguesa1 almejando três contribuições:

1. Processar e estruturar os dados produzidos pelo corpo legislativo;
2. Extrair e processar informações relevantes dentro do discurso político – em especial, com sumarização automática, modelagem de tópicos e análise de sentimentos;
3. Providenciar formas de interpretação, visualização e interação que facilite o entendimento do cidadão comum.

**Métodos e Procedimentos**

Para a criação deste sistema, foi necessário primeiramente usar métodos de *web crawling* e *web scraping* para coletar os dados das atas diretamente do *website* do Parlamento Português. Estes dados crus passam então por uma pré-segmentação de modo a dividir o texto inteiro do conjunto de Atas em uma base de dados estruturada, como mostrado na Figura 1, facilitando assim o processamento e a extração de informações a partir destes dados.

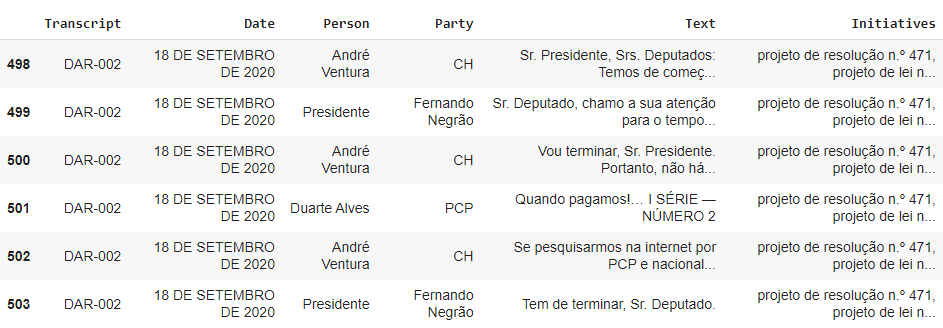


Figura 1: Dados gerados a partir das atas do DAR.

Nesses dados, vale ressaltar as colunas “Text” e “Initiatives”. A primeira corresponde a cada uma das falas que ocorreram na sessão plenária, divididas por locutor; a segunda trata de qual item de pauta corresponde aquela fala (por exemplo, sobre um Projeto de Resolução).

A partir dessa estruturação, as interações do usuário com o sistema seguem o esquema da Figura 2. O usuário insere informações que lhe são pertinentes, como o período de interesse, um partido ou político específico ou ainda um determinado assunto de interesse.

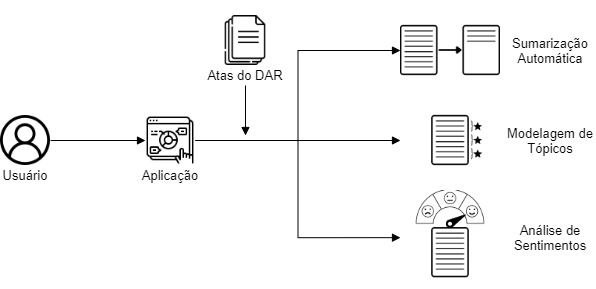


Figura 2: Esquema do sistema de Democracia Aumentada proposta, com funcionalidades de sumarização automática, definição do tópico abordado e análise dos sentimentos das falas.

Em conjunto com as informações fornecidas pelo usuário, algoritmos de AM sofisticados são aplicados, como BERT [2] e suas variantes do estado-da-arte, visando realizar três funcionalidades especializadas:

1. Sumarização Automática [3]: visa produzir um resumo a partir de um conjunto de documentos de entrada;
2. Modelagem de Tópicos [4]: método para identificar conjuntos de palavras para determinar o tópico abordado em um texto;
3. Análise de Sentimentos [5]: tarefa que almeja determinar a polaridade do sentimento expresso em um texto.

Para esse trabalho, utilizamos (1) para reduzir o tamanho das discussões de um determinado item de pauta; (2) como uma ferramenta de busca, orientando o usuário para discussões que são de seu interesse e (3) para certificar que os argumentos apresentados por um partido são condizentes com seu voto. O trabalho ainda está em andamento, com previsão de término em fevereiro de 2022.

**Resultados**

Como resultado desse projeto de iniciação científica temos um sistema unificado que engloba diversas funcionalidades de AM e PLN aplicadas no contexto das Atas do Diário da Assembleia da República Portuguesa.

Um dos resultados parciais pode ser visto na Figura 3, em que o algoritmo da modelagem de tópicos encontra as principais palavras que estão relacionadas a um mesmo assunto.

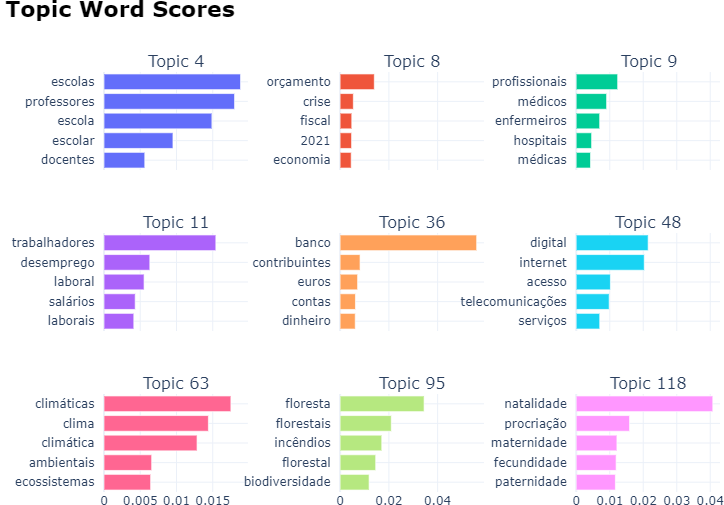


Figura 3: Exemplo dos principais tópicos identificados pelo BERTopic nas atas.

Outro resultado, agora da sumarização automática, pode ser visto na Figura 4, no qual um texto de intervenção2 de 7903 caracteres é reduzido para apenas um parágrafo sem grandes perdas no sentido da fala completa.

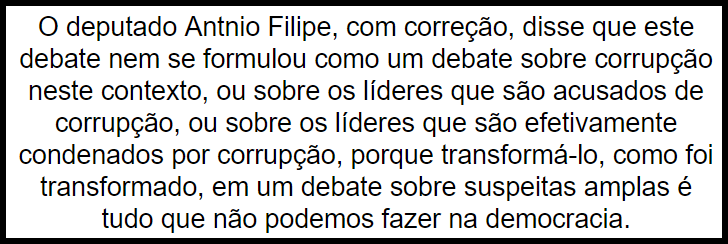


Figura 4: Exemplo de um sumário automático gerado pelo PEGASUS usando técnicas abstrativas.

**Conclusões**

O sistema de Democracia Aumentada proposto neste resumo visa contribuir principalmente para o eixo da informação da democracia digital, além das técnicas de inteligência artificial planejamos criar um aplicativo como trabalho futuro, sempre almejando uma maior qualidade para a informação recebida pelo usuário.

Além da contribuição para outros projetos que queiram se aprofundar nos demais eixos, este trabalho também contribuiu com três artigos para a área de PLN; [6] aborda técnicas de abordagem para debates moderados, [7] se aprofunda em um método de classificação após a Modelagem de Tópicos e [8] trata especificamente da ferramenta homônima a este resumo para combate à desinformação.

**Referências Bibliográficas**

*[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”.*