

## Uma Ferramenta Web para Anotações de Emoções em Textos

Alexandre S. Medeiros (IFPB, Campus Picuí), Lucas Eduardo Q. Dantas (IFPB, Campus Picuí), Gustavo de Macedo Pinto (IFPB, Campus Picuí), André Luiz Firmino Alves (IFPB, Campus Picuí)

**E-mails:** alexandre.soutomds@gmail.com, lucaseduardomdg@gmail.com, gmacedo.pinto@gmail.com, andre.alves@ifpb.edu.br

**Área de conhecimento:**(Tabela CNPq): 1.03.03.04-9 Sistemas de Informação.

**Palavras-Chave:** ferramenta web; anotação de textos; javascript; firebase.

### 1 Introdução

A análise de sentimentos tem sido uma das áreas mais pesquisadas no campo de Natural Language Processing (LIU, 2015). Descobrir o que as pessoas pensam, segundo Pang e Lee (2008), sempre foi objeto de interesse. No contexto da Web Social, através da popularização de plataformas que fornecem acesso a grande quantidade de dados subjetivos, compreender de forma automática a opinião das pessoas sobre algum tema, serviço ou produto tem sido um fator essencial para a tomada de decisões.

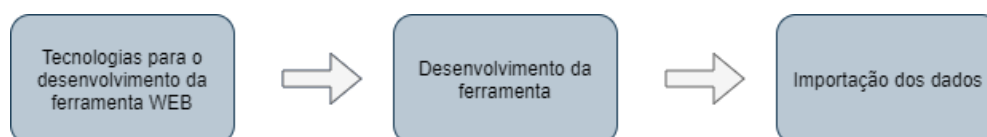
As principais abordagens utilizadas para detecção do sentimento estão divididas em quatro técnicas: aprendizagem de máquina (classificação ou regressão); análise léxica (análise linguísticas ou baseadas em dicionários), que utilizam dicionários de palavras com sentimentos já identificados; estatísticas que avaliam a co-ocorrência de termos; e semânticas, que definem a polaridade de palavras em função de sua proximidade semântica com outras de polaridade conhecidas. As diferentes técnicas podem ser combinadas para a obtenção de melhores resultados. As abordagens de aprendizagem de máquina têm sido mais exploradas nos últimos anos (KUMAR; GARG, 2020).

Nas técnicas que utilizam aprendizagem de máquina supervisionada, é necessário dispor de um conjunto de dados para realização de treinamento e teste dos modelos propostos. Estes dados devem estar classificados com as respostas esperadas (gabaritos) e geralmente, humanos participam deste processo de classificação de respostas.

O objetivo deste trabalho é apresentar uma ferramenta Web desenvolvida com tecnologias emergentes, como JavaScript, React e Banco de Dados NoSQL, que através de um processo de importação de dados textuais, possibilita a construção de dados anotados com emoções através de um processo colaborativo online, onde usuários podem ser cadastrados para rotular os dados textuais, facilitando assim o processo de construção de dados rotulados. Desta forma, esta ferramenta facilita o processo de construção de conjunto de dados para tarefas de aprendizagem de máquina supervisionada.

### 2 Materiais e Métodos

Nesta seção são descritos os materiais e métodos utilizados. A Figura 1 apresenta as principais etapas desenvolvidas neste trabalho.



**Figura 1: Etapas de desenvolvimento do trabalho.**

O primeiro passo para a criação da ferramenta era escolher as tecnologias para o desenvolvimento da ferramenta WEB. A linguagem escolhida foi *Javascript* pois é voltada para criação de páginas WEB, além de conter inúmeros frameworks e bibliotecas para facilitar todo o processo de desenvolvimento. Foi também utilizado o *React.js*, para criação de interfaces e componentes WEB.

Para a hospedagem do site e armazenamento dos dados, foi utilizado o Firebase, que é um banco de dados gratuito, online, sendo utilizado no *Back-end*.

Para desenvolvimento do *Front-end* da ferramenta Web, o primeiro passo foi realizar a criação de um protótipo da tela inicial, conforme Figura 2. Esse protótipo foi utilizado de base para criação dos componentes e estilização da ferramenta.

Figura 2: Protótipo da ferramenta

Para garantir a identificação dos usuários no sistema, foi necessário implementar um sistema de autenticação. E para evitar complicações com opções de recuperação de senhas, foi decidido utilizar a autenticação dos serviços da Google. A autenticação do usuário é feita com o email do google por meio do método *firebase.auth()*.

O site foi dividido em 3 telas: a tela de login que faz a autenticação do usuário, a tela inicial da ferramenta que exibe os dados a serem anotados, os que já foram anotados e uma tela de ajuda, em que é exibido as instruções para o usuário realizar a classificação dos textos de forma correta, e o objetivo da pesquisa.

Os dados utilizados pela ferramenta são importados e armazenados no Cloud Firestore que é um banco de dados NoSQL orientado a documentos, os quais são organizados em coleções. A coleção *Users* armazena as informações dos usuários, e a coleção *Tweets* armazena os dados a serem anotados por esses usuários.

Os dados importados para o sistema do *Firebase* foram coletados anteriormente por (LIMA et al., 2020). Esses dados eram *tweets* que estavam gravados em arquivos CSV. O primeiro passo foi criar um programa em JavaScript para ler esses arquivos. Em seguida foi utilizado o método *shuffle* para embaralhar e selecionar aleatoriamente 5000 tweets que foram enviados para o Cloud Firestore. Nesta etapa de importação, salienta-se que qualquer arquivo de texto no formato CSV proposto poderá ser utilizado para o processo de importação dos dados na ferramenta, assim como a configuração dos rótulos desejados, tornando a ferramenta genérica para qualquer processo de anotação de textos.

### 3 Resultados e Discussão

A Figura 3 apresenta visualização da página inicial da ferramenta, onde os tweets são listados de forma distinta e aleatória para cada usuário, garantindo uma maximização na coleta de dados pelos usuários. Cada *tweet* é representado por um bloco que apresenta o texto a ser analisado, e as 6 emoções (rótulos) definidas previamente (alegria, amor, raiva, medo, tristeza e ironia) que o usuário poderá classificar o texto de acordo com o seu senso crítico.

Figura 3: Tela de início da ferramenta

Portanto, se não for detectada nenhuma emoção a classificação poderá ser rotulada como nula, indicando que não há classificação para o texto. Caso se apresente divergência na classificação de um texto votado por mais de um usuário, este será classificado de acordo com a emoção predominantemente mais votada.

A ferramenta foi disponibilizada por um período de aproximadamente 1 mês para que diversos colaboradores participassem de um processo de classificação de emoções de textos dos tweets, objetivando a construção de um conjunto de dados anotados com emoções. O gráfico da Figura 4 apresenta o quantitativo de texto para cada emoção, onde cerca de 1954 tweets analisados pelos colaboradores.

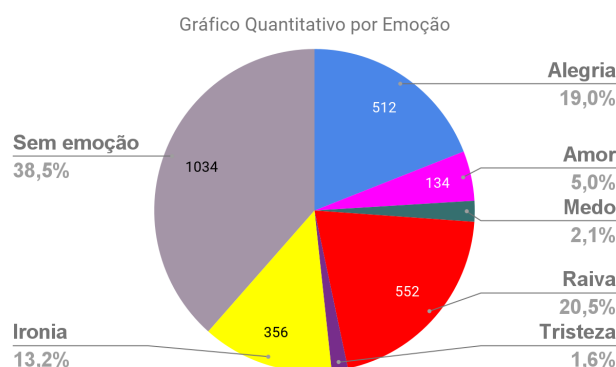


Figura 4: Quantitativo de anotações dos tweets por emoção

#### 4. Considerações Finais

Neste trabalho foi apresentado uma ferramenta WEB para anotação de emoções de textos que objetiva facilitar o processo de criação de dados anotados no idioma português para tarefas de Processamento de Linguagem Natural. O conceito da ferramenta é possibilitar a anotação colaborativa de textos, podendo configurar os rótulos desejados e a importação dos textos a serem rotulados. A ferramenta já foi utilizada e testada em um processo de classificação de emoções em texto. Deseja-se evoluir no aperfeiçoamento da ferramenta com a possibilidade de adicionar um perfil de usuário para administração da ferramenta, possibilitando bloqueio de usuários e configurações dos textos a serem importados e rótulos das anotações.

#### Agradecimentos

Os autores agradecem ao CNPq/PIBIC-EM e IFPB por financiarem esta pesquisa.

#### Referências

- LIMA, D. R. M. et al. Análise de Sentimentos em Tweets: um Estudo de Caso sobre os cortes de orçamentos nas IFEs. In: ENCONTRO NACIONAL DE COMPUTAÇÃO DOS INSTITUTOS FEDERAIS (ENCOMPIF), 7. , 2020, DOI: <<https://doi.org/10.5753/encompif.2020.11064>>.
- LIU, Bing. Sentiment Analysis: Mining Opinions, Sentiments, and Emotions. Cambridge University Press, 2015, 384p.
- PANG, Bo, LEE, Lillian, Opinion Mining and Sentiment Analysis, Foundations and Trends in Information Retrieval, Vol 2, Issue 2, 1–135, 2008. Disponível em: <<https://doi.org/10.1561/1500000011>>.
- KUMAR, A.; GARG, G. (2020). Systematic literature review on context-based sentiment analysis in social multimedia. Multimedia Tools and Applications, 79(21-22):15349– 15380.