

**Universidade Federal de Alagoas**  
**Instituto de Computação**  
Rede de Computadores 2020.E - Professor Leandro Sales

**Projeto de aplicação de sockets**

Luana Júlia Nunes Ferreira  
Lucas Mendes Massa

**Proposta da aplicação:** vetorizar um conjunto de textos e calcular coordenadas no plano para exploração de acervos por meio de gráficos, utilizando conceitos de processamento de linguagem natural (PLN).

**I. Principais funcionalidades da aplicação**

- a) CV: serve para o usuário enviar um conjunto de textos para o servidor e receber como resposta as representações vetoriais dos mesmos obtidas através do algoritmo Count Vectorizer. Esse comando retorna status 200 (SUCCESS) caso a requisição seja bem sucedida, status 300 (INVALID DATA) caso os textos de entrada não sigam os padrões necessários e status 301 (ALGORITHM RUN FAILED) caso haja alguma falha ao executar o algoritmo nos textos enviados.
- b) TFIDF: serve para o usuário enviar um conjunto de textos para o servidor e receber como resposta as representações vetoriais dos mesmos obtidas através do algoritmo TF/IDF. Esse comando retorna status 200 (SUCCESS) caso a requisição seja bem sucedida, status 300 (INVALID DATA) caso os textos de entrada não sigam os padrões necessários e status 301 (ALGORITHM RUN FAILED) caso haja alguma falha ao executar o algoritmo nos textos enviados.
- c) DISCONNECT: serve para o usuário encerrar a conexão com o servidor e finalizar a execução do cliente. Esse comando retorna status 500 (DISCONNECTED) caso a requisição de desconexão seja bem sucedida. OBS.: Se o comando digitado for inválido, a resposta do status será 400 (INVALID COMMAND).

**II. Sugestões para a continuação do projeto**

- a) Poderia ser implementado um cabeçalho com o fito de permitir que os usuários passem parâmetros aos algoritmos de vetorização.
- b) Poderia ser implementada uma forma que permitisse ao protocolo lidar com o tamanho variável dos textos de entrada.
- c) Poderia ser implementada uma função para que o protocolo pudesse manipular diferentes tipos de entrada.

**III. Dificuldades durante o desenvolvimento**

- a) Garantir o formato da mensagem do protocolo.
- b) Lidar com arquivos maiores.
- c) Validar o modelo em 100%, visto que os testes de software aplicados, ainda não há garantia de que não há falhar.