



FUNDAÇÃO EDSON QUEIROZ
UNIVERSIDADE DE FORTALEZA
Programa de Pós-Graduação em Informática Aplicada –
PPGIA

TRABALHO FINAL DA DISCIPLINA – TÓPICOS AVANÇADOS EM PLN – 2021.2

Objetivos:

Desenvolver o módulo de NLU (*Natural Language Understanding*) de um sistema para chatbot fim-a-fim que recebe dois textos em linguagem natural e define se o 2º texto é resposta para o 1º texto.

Será fornecido um dataset com 180000 pares de sentenças que representam **diálogos positivos**, em inglês. Os diálogos negativos deverão ser “montados” pela equipe.

As equipes serão avaliadas pelos entregáveis e pelos resultados a partir de um conjunto de teste (dados não vistos – Held-Out Data) com 18000 pares de sentenças.

Produtos Entregáveis:

- Pré-processamento, Lematização e POS Tagger, NER, etc. –
18/10/2021
- Estatística Descritiva do vocabulário constante no dataset, considerando os modelos de linguagem unigrama, bigrama, trigrama; análise da lei de potencia ou Lei de Zipf –
18/10/2021
- Vetorização das Sentenças (*features*) – **18/10/2021**
- Model Fitted e Resultados do processo de treinamento -
06/12/2021
- Execução com Held-Out-Data (Dados não vistos) –
06/12/2021
- Artigo resumo com 4 paginas - **06/12/2021**

Apresentação Oral e Entrega do artigo: 06/12/2021