

Trabalho Prático 6 - Processamento de Linguagem Natural (PLN)

Objetivos

Consiste em praticar os conceitos introdutórios de processamento de linguagem natural vistos em sala de aula.

Descrição

1. Acesse a base de dados disponibilizada (no GitHub), conforme exemplo abaixo;

```
1 import requests
2 req = requests.get("https://raw.githubusercontent.com/juliosoareshreis/NLP/main/data/biblia.
  txt")
3 url_content = req.content
4 arquivo = open("biblia.txt", "wb")
5 arquivo.write(url_content)
6 arquivo.close()
```

2. Analise a base de dados e plote os unigramas e bigramas mais comuns;
3. Faça uma vetorização com `Word2vec` ou `FastText` – vetor de 64 posições;
4. Selecione 10 palavras e para cada uma delas faça uma lista de 5 palavras mais similares;
5. Faça a análise PCA e plote as palavras do item anterior.

Entrega

A entrega deve ser efetuada conforme agendado no PVANet Moodle. Envie APENAS UM arquivo contendo a sua implementação. Caso não consiga submeter o arquivo (Colab) diretamente, submeta um arquivo contendo o referido *link*.

Código: <https://colab.research.google.com/drive/1ulKQ2J31xzQTpnGx4O9MEde3TGptikF?usp=sharing>

Comentários Gerais

- O trabalho deve ser realizado individualmente (grupo de UM aluno);
- Trabalhos copiados serão penalizados (NOTA Zero).