

Trabalho - v1

Instruções:

- O trabalho deve ser feito em trios.
- Apenas 1 integrante do trio deve entregar os arquivos, mas todos devem responder o questionário.
- **Não utilize nenhuma biblioteca, todo o código usado deve ter sido desenvolvido pela equipe.**
- **Não será aceito entregas atrasadas.**
- Relatório:
 - Só será aceito em PDF.
 - Deve ter no máximo 4 páginas e deve ser no formato de artigo com coluna dupla.
 - Deve conter partes do código, diagramas e as explicações de como ele foi feito, exemplos de resultados, definições e escolhas feitas pela equipe.
 - Deve conter uma conclusão com a opinião do grupo sobre os resultados.
- A entrega será em um Google Forms.
- Caso um dos arquivos não seja entregue, a nota da equipe será zerada
- O que **deve** ser entregue:
 - Código fonte da aplicação;
 - Relatório em PDF;
 - Arquivos de entrada;
 - Arquivo de saída gerado pela aplicação de vocês com os arquivos de entrada.

Critérios de avaliação:

- Relatório (40 pontos):
 - Legibilidade
 - Descrição do problema
 - Apresentação de resultados
 - Discussão dos resultados
 - Código (60 pontos)
 - Legibilidade do código (10 pontos)
 - Organização do código (20 pontos)
 - utilização de funções
 - códigos em arquivos distintos na mesma pasta (usar import)
 - Tratamento de erros (10 pontos)
 - Try-Except
 - Tratamento de leitura de arquivos inválidos
 - Arquivos (20 pontos)
 - Leitura dos arquivos txt
 - Escrita do arquivo CSV
 - Padronização correta do arquivo CSV
-

Contexto:

Análise de sentimentos é uma área da computação que visa interpretar o sentimento de uma pessoa com base em um texto, existem diversas formas de se fazer isso, as mais atuais e utilizadas utilizam Deep Learning, envolvendo técnicas de processamento de linguagem natural (NLP), análise de texto, biométricas (Métricas que tentam calcular ou relacionar as relações humanas), entre outras. Essas ferramentas são muito uteis e empregadas em diversas áreas, como: investimentos em bolsa de valores, classificação de filmes, combate a cyber bullying, combate a pedofilia entre outras. Neste trabalho, iremos desenvolver uma forma simplificada de análise de sentimentos.

Mais detalhes a boa e velha wikipedia:

https://en.wikipedia.org/wiki/Sentiment_analysis

A empresa WalID está fazendo um concurso para escolher programadores que irão trabalhar em sua mais nova equipe de TI. No concurso, os grupos devem desenvolver uma das primeiras formas utilizadas para a análise de sentimento, onde irão detectar o quão ruim ou boa a frase é, sendo capaz de “aprender” novas palavras. A aplicação deve ser capaz de:

1. Ler **um** arquivo no formato .txt contendo as palavras e pesos de referência:
 - O nome do arquivo deve ser definido pela equipe;
 - A organização do arquivo deve ser definido pela equipe;
2. Ler **um** arquivo no formato .txt contendo as frases que serão analisadas.
 - O nome do arquivo deve ser definido pela equipe;
 - A organização do arquivo deve ser definido pela equipe;
3. Ao fim, a aplicação deve salvar **um** arquivo no formato .csv (Comma-separated values, valores separados por vírgulas) com cabeçalho e contendo:

- Frase;
- Lista das palavras reconhecidas;
- Resultado do cálculo de sentimento;
- As palavras que foram “aprendidas”.

Para isso a equipe deve seguir criteriosamente os seguintes passos:

- Procurar um conjunto de frases que serão usadas para validar o algoritmo. Escolham no mínimo 10 frases, **cada membro da equipe** (individualmente) deve classificar as frases na escala -5 a 5 (-5 muito ruim até 5 muito boa) conforme sua interpretação.
 - a. É esperado que no relatório tenha uma boa comparação entre o que cada membro da equipe classificou a frase e o que o algoritmo classificou.
 - b. Dica: Escolha frases simples e curtas (+-140 caracteres), não estamos usando um algoritmo robusto e não será capaz de interpretar ironias (ou será que vai?). Frases de efeito de filmes e séries ou twitters podem ser uma boa escolha.
- Criar manualmente **um arquivo no formato .txt** de palavras e atribuir pesos para elas. Esses pesos costumam estar relacionados com o sentimento da palavra. Crie uma lista com 30 palavras, e atribua pesos para elas em uma escala de -5 a 5, sendo: -5 frases muito ruins e 5 frases muito boas.

Ex: Palavras e pesos

Palavra	Peso
Amigo	3
Excelente	5
Não	-1
Guerra	-5

- Criar manualmente **um arquivo no formato .txt** contendo as frases que serão analisadas.
- Desenvolver um algoritmo que leia as frases no arquivo txt e salve **um arquivo no formato .csv** como descrito no enunciado (item 3) (Os pesos das frases não podem ultrapassar o range -5 até 5).
- A classificação da frase depende unicamente da média dos pesos dela, dessa forma uma frase que a média é -5 é considerada muito ruim, já uma frase com média 0 é considerada neutra.
- A aplicação deve “aprender” palavras novas, para isso quando uma palavra não listada aparecer, a aplicação deve atualizar o arquivo de palavras e pesos, adicionando essa nova palavra e atribuindo o peso dela com o peso dado à frase. (Os pesos das palavras não podem ultrapassar o range -5 até 5)