

Classificação com bert4re - Tutorial

Visão rápida

Sobre o bert4re: 'bert4re' é um modelo baseado em no modelo BERTbase treinado para textos de engenharia de software. Foi utilizado uma pipeline do tipo "text-classification" pois ele não é um modelo de Inferência de Linguagem Natural(NLI).

Este script:

- Varre todo o código-fonte do ChatTTS
- Usa o modelo **BERT4RE** para detectar padrões arquiteturais
- Gera um arquivo **CSV** com os 3 padrões mais prováveis por arquivo

Útil para engenharia reversa, documentação e pesquisa.

Pré-requisitos (Arquivo ".ipynb")

Navegador Web e conta no google

Execução

1) Abra o arquivo ".ipynb" no Google Colab

Selecione e execute todas as células.

Confira o arquivo ".csv" gerado.

Pré-requisitos (Arquivo ".py")

No terminal:

```
pip install transformers torch accelerate tqdm pandas
```

Execução

1) Baixe o ChatTTS

```
git clone https://github.com/2noise/ChatTTS.git
```

2) Rode o script

```
python classify_chatts_architecture.py --repo ChatTTS
```

Arquivo Gerado

O script criará:

`chatts_architecture_labels.csv`

Formato:

file	top_labels
ChatTTS/core/encoder.py	model, pipeline, repository
ChatTTS/utils/audio.py	pipeline, layers, client-server