

Para rodar o projeto:

1. Acessar a pasta '/correferencia'
2. No terminal executar: **'python3 -m venv .env'** (para setar um ambiente virtual local)
3. No terminal executar: **'source .env/bin/activate'** (nesta pasta o python default sera 3.x)
4. No terminal executar: **'pip install -U spacy==2.1.0'** (versão mais atual dá problema com a biblioteca neuralcoref)
5. No terminal executar: **'pip install -U pandas'**
6. No terminal executar: **'python -m spacy download en\_core\_web\_sm'**
7. No terminal executar: **'pip install neuralcoref --no-binary neuralcoref'**
8. No terminal executar: **'pip install googletrans'** (instável, pode ser que dê problema de quantidade de requisições)
9. No terminal executar: **'python correfer.py'** (rodar o código em si)

A separação das sentenças da entrada foi feita no arquivo "**process\_csv.py**".  
Escrevemos o arquivo **sentences.txt** com cada sentença em cada linha.

A extração de correferência foi feita com a biblioteca Spacy e Neuralcoref.

A entrada em português para o *neuralcoref* não funciona muito bem portanto traduzimos as entradas para inglês antes e então realizamos a detecção de correferência.

A tradução do arquivo de entrada foi feita manualmente no site do Google Translate e o resultado foi manualmente copiado para o arquivo **translated\_sentences.txt**

O script **correfer.py** processa este arquivo e escreve o resultado obtido no arquivo **results.txt**

O arquivo **results.txt** foi traduzido manualmente assim como as sentenças o que pode gerar uma inconsistência nas traduções originais. Poderia ter sido resolvido com a árvore sintática porém as APIs de tradução tinham limitações de requisição o que tornava mais complexo essa implementação.

Houve uma dificuldade na instalação da biblioteca até descobrir que havia problemas de compatibilidade com versões recentes do spaCy.

Também houve dificuldade de saber métodos para tratar os casos em que o neuralcoref não conseguia uma resposta ou dava uma resposta incompleta.