SCC0633 - Processamento de Linguagens Naturais

Anotação do Córpus

Elisa Saltori Trujillo, 8551100

Laura Cruz Quispe, 10556941

Leonardo Cesar Cerqueira, 8937483

Objetivo

- → Análise de sentimento de tweets
 - Polaridade positiva ou negativa



- → Tema: Novela O Outro Lado do Paraíso
 - Programa popular
 - Opiniões controversas



Coleta de Tweets

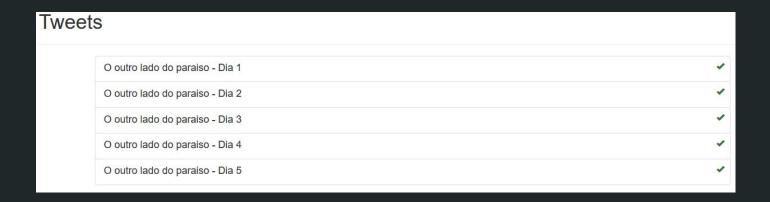
- → Feita em Python
 - ◆ API <u>Tweepy</u>
- → 1000 Tweets escolhidos para anotação
 - Distribuídos em 5 dias de novela
 - Selecionados aleatoriamente
 - Evitar viés

Categorias

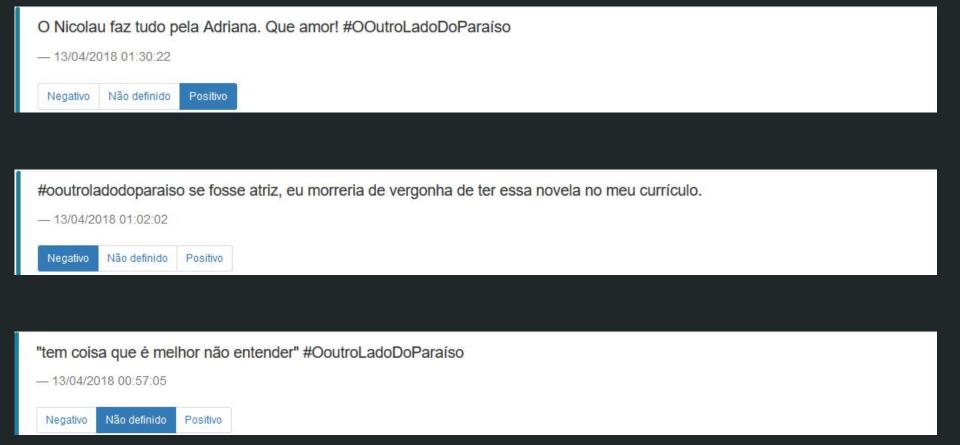
- → Positivo
 - Sentimentos positivos sobre a novela
- → Negativo
 - Sentimentos negativos sobre a novela
- → Não definido
 - Tweets ambíguos, mal estruturados. Serão descartados
- → Aplicados à história, atores, autor, etc.
- Ferramenta para auxiliar na interpretação de gírias: Dicionário inFormal

Ferramenta de anotação

- → Baseada na aplicação <u>ScoreSite</u>
 - Simples, fácil de usar
- → Tweets divididos em 5 dias
- → Três Botões
 - Negativo, Não Definido, Positivo



Exemplos



Resultados Medidas de Concordância

→ Simples

Anotadores	% de Concordância
Laura e Leonardo	0.599
Laura e Elisa	0.61
Leonardo e Elisa	0.778
Todos	0.51

→ Kappa

Anotadores	Coeficiente Kappa
Laura e Leonardo	0.4135
Laura e Elisa	0.4066
Leonardo e Elisa	0.6614

- → Composição final: 659 tweets válidos
 - 247 positivos
 - 412 negativos

Dúvidas?