

# SCC0633 - Processamento de Linguagens Naturais

---

## Anotação do Corpus

---

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 Tweepy

→ 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 ScoreSite
  - ◆ Simples, fácil de usar
- Tweets divididos em 5 dias
- Três Botões
  - ◆ Negativo, Não Definido, Positivo

Tweets	
O outro lado do paraíso - Dia 1	✓
O outro lado do paraíso - Dia 2	✓
O outro lado do paraíso - Dia 3	✓
O outro lado do paraíso - Dia 4	✓
O outro lado do paraíso - Dia 5	✓

# Exemplos

O Nicolau faz tudo pela Adriana. Que amor! #OOutroLadoDoParaíso

— 13/04/2018 01:30:22

Negativo

Não definido

Positivo

#outroladodoparaíso se fosse atriz, eu morreria de vergonha de ter essa novela no meu currículo.

— 13/04/2018 01:02:02

Negativo

Não definido

Positivo

"tem coisa que é melhor não entender" #OoutroLadoDoParaíso

— 13/04/2018 00:57:05

Negativo

Não definido

Positivo

# 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?