

## Trabalho 1 - Graduação

Acesse a equipe do teams, em material de aula.

Escolha um dos seguintes datasets:

- Hatespeech-BR
- MQD-1465
- Youtube-Depress
- PREDADORES-BR

E um dos seguintes léxicos:

- Netspeak
- AffectPT-BR
- MDF-BR
- Depress-pt-BR

PS: o pareamento do artigo x produto (léxico ou dataset) se encontra ao final deste documento.

O trabalho deve ser feito em linguagem JAVA. NÃO PODE usar biblioteca.

Passo 1 - Cada aluno precisa escolher uma combinação de 1 dataset e 1 léxico. Não pode ser duplicado. Um dataset pode ser escolhido por no máximo dois alunos e um léxico por no máximo dois alunos. Exemplo: Luana escolheu Hatespeech-BR e Netspeak. Juliana pode escolher Hatespeech-BR mas não netspeak. Mariana não pode escolher Hatespeech-BR, pois dois alunos já escolheram. Colocarem no Teams no canal Geral publicamente.

Passo 2 - Exibir as 100 palavras que mais ocorrem no dataset escolhido e a quantidade de vezes que ela ocorreu. Ordem decrescente. Ex:

casa - 1000

bola - 134

mesa - 18

...

Passo 3 - Com relação ao léxico. Conte quantas vezes aparece cada palavra do léxico no conjunto de dados escolhido. Exiba as 20 palavras que mais ocorrem com sua respectiva contagem. Ex: léxico do Netspeak

vc - 345

tb - 89

tbm - 43

...

Passo 4 - Apresentar o artigo em 15 min (cada um escolhe um artigo diferente). Colocarem no Teams no canal Geral publicamente ([o nome do artigo](#)).

Passo 5 - Entregar **apenas** uma seção de resultados, escrito em latex conforme modelo do artigo apresentado. (sugestão, usar o overleaf). Exemplo de seção de resultados (não precisa ser nada muito elaborado):  
<https://sol.sbc.org.br/index.php/brasnam/article/view/16137>

Apêndice I

***Pareamento artigo (na pasta artigos) X produto (na pasta léxico ou dataset)***

Identificando Sinais de Comportamento Depressivo em Redes Sociais - (**Depress-pt-BR - léxico**)

Identificando sintomas depressivos- um estudo de caso no Youtube (**Youtube-Depress - dataset**)

Netspeak-BR - Um lexico sobre expressoes criadas na lingua portuguesa brasileira para a Internet (**Netspeak - léxico**)

Hate Speech Detection Using Brazilian Imageboards (**Hatespeech-BR - dataset**)

NAT Towards an Emotional Agent - (**MQD-1465 - dataset**)

Identificação de predadores sexuais brasileiros por meio de análise de conversas realizadas na Internet - (**PREDADORES-BR - dataset**)

A Brazilian Portuguese Moral Foundations dictionary (**MDF-BR - léxico**)

AffectPT br an Affective Lexicon based on LIWC 2015 (**AffectPT-br - léxico**)