

**CENTRO UNIVERSITÁRIO CARIOCA**

**WASHINGTON LUIZ SOBRAL MENÊZES**

**PROCESSAMENTO DE LINGUAGEM NATURAL (PLN), UTILIZANDO O PYTHON COM A BIBLIOTECA NLTK.**

**RIO DE JANEIRO**

**2022**

**WASHINGTON LUIZ SOBRAL MENÊZES**

**PROCESSAMENTO DE LINGUAGEM NATURAL (PLN), UTILIZANDO O PYTHON COM A BIBLIOTECA NLTK.**

Trabalho de Conclusão de Curso apresentado ao Centro Universitário Carioca, como requisito parcial para obtenção do grau de Bacharel em Engenharia da Computação.

Orientador: Prof. D.Sc Sérgio Assunção Monteiro

**RIO DE JANEIRO**

**2022**

**WASHINGTON LUIZ SOBRAL MENÊZES**

**PROCESSAMENTO DE LINGUAGEM NATURAL (PLN), UTILIZANDO O PYTHON COM A BIBLIOTECA NLTK.**

Trabalho de Conclusão de Curso apresentado ao Centro Universitário Carioca, como requisito parcial para obtenção do grau de Bacharel em Engenharia da Computação.

BANCA EXAMINADORA

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Prof. Sérgio Monteiro Assunção, D.Sc - Orientador

Centro Universitário Carioca

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Prof. André Luiz Avelino Sobral, M.Sc

Centro Universitário Carioca

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Prof. Alberto Tavares da Silva, D.Sc

Centro Universitário Carioca

**RESUMO**

Este artigo tem por objetivo relatar a iniciativa e inovação dos recursos da tecnologia de processamento de linguagem natural utilizando a linguagem Python com pacotes de biblioteca NLTK, que se concentra em tornar a linguagem humana natural utilizável por programas de computador e verificar o seu comportamento, sua utilização, finalidades, e desempenho para, posteriormente, propor em melhor entendimento dos pacotes NLTK.

**Palavras-chave:** Python, NLP, NLTK;

**ABSTRACT**

This article aims to report the initiative and innovation of natural language processing technology resources using the Python language with NLTK library packages, which focuses on making natural human language usable by computer programs and verifying its behavior, its use, purposes, and performance to later propose a better understanding of NLTK packages.

**Keywords:** Python, NLP, NLTK;

**LISTA DE ILUSTRAÇÕES**

[Figura 1 - Python e NLTK, linguagem natural, processamento e analise 1](#_Toc467795380)0

[Figura 2 - Sentenças da Tokenização](#_Toc467795381) 14

[Figura 3 - Fórmula de Naive Bayes](#_Toc467795382) 15

[Figura 4 - Site do Python, Fonte autor](#_Toc467795383) 19

[Figura 5 - Foto do Instalador, Fonte autor 2](#_Toc467795384)0

[Figura 6 - Foto do Instalador, Fonte autor 2](#_Toc467795385)0

[Figura 7 -](#_Toc467795386) [Foto do Instalador, Fonte autor](#_Toc467795385) [2](#_Toc467795386)1

[Figura 8 - Foto do Prompt de Comando, Fonte autor 22](#_Toc467795387)

[Figura 9 - Foto do Instalador VsCode - Fonte autor 2](#_Toc467795388)2

[Figura 10 - Foto do VsCode - Fonte autor](#_Toc467795389) 23

[Figura 11 - Foto do NLTK Download - Fonte autor 24](#_Toc467795390)

[Figura 12 - Importação de dependências - Fonte autor](#_Toc467795391) 25

[Figura 13 - Implementação do Tokenize - Fonte autor](#_Toc467795392) 26

[Figura 14 - Implementação do Stemming - Fonte autor](#_Toc467795393) 26

[Figura 15 - Implementação do buscapalavra - Fonte autor](#_Toc467795394) 27

[Figura 16 - Implementação do buscafrequencia - Fonte autor](#_Toc467795395) 28

[Figura 17 - Implementação do ExtratorPalavrasTreinamento - Fonte autor](#_Toc467795396) 29

[Figura 18 - Implementação do Classificador - Fonte autor](#_Toc467795397) 29

[Figura 19 - Implementação da matriz de confusão - Fonte autor](#_Toc467795398) 30

[Figura 20 - Implementação do teste de Stemming - Fonte autor](#_Toc467795399) 31

[Figura 21 - Implementação da impressão dos testes - Fonte autor](#_Toc467795400) 32

[Figura 22 - Implementação da impressão Tokenize- Fonte autor](#_Toc467795401) 32

[Figura 23 - Implementação do Teste.py - Fonte autor](#_Toc467795402) 33

[Figura 24 - Resultado da acuracidade BaseCompletaTreinamento - Fonte autor](#_Toc467795403) 34

[Figura 25 - Resultado da Matriz de confusão da Base - Fonte autor](#_Toc467795404) 34

[Figura 26 - Resultado de “odeio ir a escola” - Fonte autor](#_Toc467795404) 35

[Figura 27 - Resultado de “*gostamos de ir a escola”* - Fonte autor](#_Toc467795404) 36

[Figura 28 - Resultado de *“não acredito que amanha vai fazer sol”*- Fonte autor](#_Toc467795404) 37

SUMÁRIO

[1 INTRODUÇÃO](#_Toc467795622) 09

[1.1 Objetivo do Trabalho](#_Toc467795625) 09

[1.2 Estrutura do Trabalho 1](#_Toc467795625)0

[2 REVISÃO BIBLIOGRÁFICA 1](#_Toc467795623)1

[2.1 Historia do Python e da biblioteca NLTK 1](#_Toc467795625)1

[2.2 Os princípios do NLTK 1](#_Toc467795625)2

[2.2.1 Simplicidade](#_Toc467795626) 12

[2.2.2 Consistência](#_Toc467795626) 12

[2.2.3 Extensibilidade](#_Toc467795626) 12

[2.2.4 Modularidade](#_Toc467795626) 12

[2.3 Ferramentas do NLTK 1](#_Toc467795625)3

[2.3.1 Stemming](#_Toc467795626) 13

[2.3.2 Stopwords](#_Toc467795626) 13

[2.3.3 Tokenização](#_Toc467795626) 14

[2.3.4 Classificação](#_Toc467795626) 15

[2.3.5 Analise](#_Toc467795626) 15

[2.4 Estudo das emoções 1](#_Toc467795625)6

[2.4.1 Surpresa](#_Toc467795626) 16

[2.4.2 Alegria](#_Toc467795626) 16

[2.4.3 Tristeza](#_Toc467795626) 17

[2.4.4 Medo](#_Toc467795626) 17

[2.4.5 Desgosto ou nojo](#_Toc467795626) 17

[2.4.6 Raiva](#_Toc467795626) 18

[3 REQUISITOS E DESENVOLVIMENTO DA METODOLOGIA DO TRABALHO 19](#_Toc467795623)

[3.1.1 Primeiro passo a instalação do Python](#_Toc467795625) 19

[3.1.2 Segundo passo Instalação da IDE e dependências.](#_Toc467795625) 22

3[.2.1 Construção da Aplicação.](#_Toc467795626) 24

[3.2.2 Construção do teste.](#_Toc467795626) 33

[4 RESULTADO DOS TESTES 3](#_Toc467795623)3

[4.1 Resultado “odeio ir a escola” 3](#_Toc467795625)5

[4.2 Resultado “gostamos de ir a escola”](#_Toc467795625) 36

[4.3 Resultado “não acredito que amanha vai fazer sol”](#_Toc467795625) 37

[5 CONSIDERAÇÕES FINAIS](#_Toc467795623) 38

[REFERÊNCIAS BIBLIOGRÁFICAS](#_Toc467795649) 39

**AGRADECIMENTOS**40

LISTA DO BANCO DE DADOS DA BASE.PY:41

LISTA DE STOPWORKS:64

**1 INTRODUÇÃO**

O Processamento de Linguagem Natural (PLN) tem como objetivo extrair representações e significados mais completos de textos livres escritos em linguagem natural, que é utilizada para fins de comunicações de uso comum por humanos; em idiomas como Português e Inglês, utilizando conceitos linguísticos como classes de palavras (adjetivo, substantivo, verbo, etc.), e estruturas gramaticais.

O Processamento de Linguagem Natural também lida com situações mais complexas, como anáforas e ambiguidades. Isso se dá através de várias representações de conhecimento, como léxicos de palavras e seus significados considerando a estrutura hierárquica da linguagem, analisando a linguagem pelo seu significado.

Os sistemas de Processamento de Linguagem Natural podem ser utilizados em uma diversa gama de papéis, como corretores gramaticais, na conversão de fala para texto, na tradução automática de texto entre linguagens e na análise de sentimentos/mineração de opiniões.

O Processamento de Linguagem Natural, tem cuja importância nas áreas científicas, econômicas, sociais e culturais e está experimentando um rápido crescimento à medida que suas teorias e métodos são implantados em uma variedade de novas tecnologias de linguagem como na linguagem Python com a biblioteca NLTK.

**1.1 Objetivo do Trabalho**

O objetivo central do trabalho é desenvolver uma aplicação, utilizando a biblioteca NLTK, para ler em uma base os dados e analisar os resultados dos testes, Observando o comportamento do Stemming, Stopwords, Tokenização, Classificação ( Naive Bayes) e Analise dos sentimento do teste aplicado.

**1.2 Estrutura do Trabalho**

O trabalho está dividido em cinco capítulos. O primeiro capítulo é introdutório e apresenta a divisão do trabalho.

O segundo capítulo começa apresentando a história do Python e da biblioteca NLTK e suas ferramentas, Após este trecho, o capítulo introduz alguns conceitos estudo das emoções.

O terceiro capítulo contém os requisitos e objetivo do desenvolvimento dos códigos. Em seguida, a estrutura de todos os códigos desenvolvidos é detalhada e justificada.

O quarto capítulo apresenta um estudo de caso dos testes das frases “odeio ir a escola”, “gostamos de ir a escola” e “não acredito que amanha vai fazer sol”, comparando os resultados das análises com as emoções abordadas. E testando a acuracidade do algoritmo. Finalizando, o capítulo 5 descreve a conclusão do trabalho.

****

Figura 1 – Python e NLTK envolvendo linguagem natural, processamento e analise.

**2 REVISÃO BIBLIOGRÁFICA**

A revisão bibliográfica será abordada sobre a história do Python e da sua biblioteca NLTK com seus princípios e ferramentas, e também será revisado a abordagem sobre os estudos das emoções.

**2.1 Historia do Python e da biblioteca NLTK**

Python foi desenvolvido em 1989, no centro de Matemática e Tecnologia da Informação (CWI, Centrum Wiskunde e Informatica), na Holanda por Guido van Rossum, em 1991 Guido publicou o código (nomeado versão 0.9.0) no grupo de discussão alt.sources (newsgroup), e foi feita a abertura para a programação.

Python foi desenvolvida para ser uma linguagem de programação de alto nível, interpretada de script, imperativa, orientada a objetos, funcional, de tipagem dinâmica, simples e forte de excelente funcionalidade para o processamento de dados linguísticos , com uma sintaxe clara, precisa e reduzida, que auxiliar a legibilidade do código fonte.

O NLTK - Natural Language Toolkit foi desenvolvido em 2001, como objetivo de uma disciplina em linguística computacional no Department of Computer and Information Science da University of Pennsylvania. E vem sido desenvolvido e acrescido com a ajuda de dezenas de pessoas.

O NLTK foi criado como uma suíte de aplicativos e módulos de código-fonte aberto denominada como biblioteca, para prover o aprendizado da linguagem natural.

**2.2 Os princípios do NLTK**

O NLTK foi desenvolvido e pensado com quatro objetivos primários em mente sendo eles simplicidade, consistência, extensibilidade e modularidade.

**2.2.1 Simplicidade**

Disponibilizar um framework intuitivo em conjunto a substanciais blocos de construção, provendo aos usuários de um conhecimento prático de NLP, sem dificultar as tarefas associadas ao processamento de dados linguísticos anotados.

**2.2.2 Consistência**

Disponibilizar um framework unificado com interfaces e estruturas de dados consistentes, e nomes de método facilmente prognosticáveis.

**2.2.3 Extensibilidade**

Disponibilizar uma estrutura no objetivo em que novos módulos de software possam ser utilizados de forma simples e fácil, adicionando implementações alternativas a abordagens variadas para uma mesma tarefa.

**2.2.4 Modularidade**

Disponibilizar componentes (módulos) que possam ser utilizados independentemente sem a necessidade de compreender o restante da ferramenta.

**2.3 Ferramentas do NLTK**

O NLTK e um kit de ferramentas para desenvolver programas em Python que trabalham com dados da linguagem humana para aplicação no processamento estatístico de linguagem natural. que possuem bibliotecas de processamento de texto. Sendo elas: Stemming, Stopwords, Tokenização, Classificação ( Naive Bayes) e Analise.

**2.3.1 *Stemming***

O Stemming é realizada uma técnica de redução de termos a um radical comum, no ponto inicial de uma análise das características gramaticais dos elementos, como partículas localizadas no final das palavras com finalidade de indicar as flexões de gênero, número dos nomes e as flexões de número, gênero, pessoa, tempo e modo dos verbos. Permitindo que você se concentre-se no significado básico de uma palavra ao contrário de todos os detalhes de como ela está sendo utilizada, reduzindo as palavras agrupando-as por meio de seu sufixo.

**2.3.2 *Stopwords***

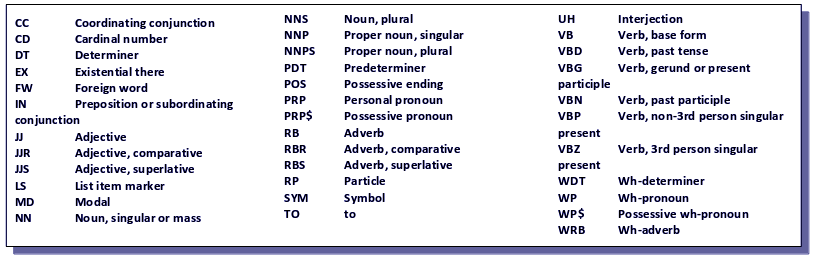
O NLTK obtém uma lista de Stopwords para o Português e Inglês, que são palavras que podem ser consideradas desnecessárias para o entendimento do sentido de um texto, ou seja, palavras semanticamente irrelevantes. Essas palavras são na maioria dos casos são removidas de um texto durante a fase de pré-processamento realizando uma filtragem e limpeza dos textos, pois são palavras comuns que não gera significado para o resultado final, possibilitando ter uma base de dados de melhor estruturada, tornando as tarefas de mineração de dados mais fácil e ágil.

**2.3.3 *Tokenização***

A Tokenização é bem estabelecida e bem entendida para linguagens artificias tais como linguagens de programação, tem como finalidade a segmentação de palavras, que é a quebra a sequência de caracteres em um texto localizando onde uma palavra ou por frases terminam e outra começa, possibilitando trabalhar com textos menores que ainda são relativamente coerentes e significativos, mesmo estando do contexto do restante do texto.

A Tokenização em linguagens delimitadas por espaço, a maioria das ambiguidades existem entre o uso de sinais de pontuação, sendo tais como: ponto final, vírgulas, aspas, apóstrofos e hifens, possibilitando ser utilizado em outras funções de mesma sentenças o mesmo sinal. Entretanto, não são necessariamente os tokens requiridos para processamento adicional, por causa de sua natureza ambígua dos sistemas de escrita, como vastas técnicas de tokenização requeridas por diversas aplicações.

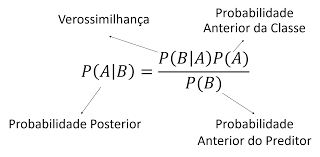
A Tokenização das linguagens não segmentadas, obrigatoriamente tem informação léxica e morfológica adicional, possuem em maioria das ambiguidades existente, entre o uso de sinais de pontuação, sendo tais como: ponto final, vírgulas, aspas, apóstrofos e hifens, sendo usado o mesmo sinal de pontuação em diferentes funções de uma mesma sentença.

Figura 2 – Sentenças da Tokenização

fonte: https://data-flair.training/blogs/nltk-python-tutorial/

**2.3.4 *Classificação***

O classificador é um algoritmo baseado em Naive Bayes que faz uma abordagem probabilística que utiliza a probabilidade baseado na regra do Teorema de Bayes assumindo independência nos atributos do objeto. Possui uma parte matemática relativamente simples, possui um bom desempenho e precisa de poucas observações para ter uma boa acurácia, Ele é frequentemente utilizado como base na classificação de textos por ser rápido e fácil de implementar.

Figura 3 – Fórmula de Naive Bayes

**2.3.4 *Analise***

A Análise de sentimento tem como seu relacionamento ao processo de análise de informação que realiza uma varredura procurando reduzir o espaço de busca, recuperando apenas as informações que são relativamente importantes para a resolução de problemas determinados, a análise realiza a validação da eficiência ou acurácia do processo como um todo, analisando os dados obtidos após aplicação dos algoritmos na etapa anterior, possuindo alguns tipos de análises, entre elas estão a análise morfológica, sintática, semântica e léxica.

Na Análise Morfológica tem como finalidade ser responsável por definir artigos, substantivos, verbos e adjetivos, armazenados em um tipo de dicionário.

Na Análise Sintática o utiliza o dicionário procurando por mostrar relacionamento entre as palavras e, em seguida, verifica sujeito, predicado, complementos nominais e verbais, adjuntos e apostos.

Na Análise Semântica, há o encontro de termos ambíguos, de sufixos e afixos, entretanto, em questões de significado associados aos morfemas componentes de uma palavra, o sentido real da frase ou palavra,

Na Análise Léxica faz converter uma sequência de caracteres em um seguimento de palavras que foram as palavras candidatas a serem termos do índice e colocando em ordem alfabética de caracteres de palavras (a-z) e separadores de palavras.

Na Análise Semântica tem como finalidade procurar e identificar a função que determinados termos exercem no texto. *(as palavras que não tem valor semântico em base textual são as Stopwords)*. podem ocorre o encontro de termos ambíguos, de sufixos e afixos, entretanto, em questões de significado associados aos morfemas componentes de uma palavra, o sentido real da frase ou palavra.

**2.4 Estudo das emoções**

Um dos objetivos deste trabalho, é detectar as emoções que podem ser interpretadas através de textos, com isso é necessário saber sobre que tipo de emoções são essas, segundo o psicólogo Paul Ekman, afirma que as principais emoções básicas são: Surpresa, Alegria, Tristeza. Medo, Desgosto ou nojo, Raiva.

**2.4.1 *Surpresa***

A surpresa é uma das emoções mais rápida e passageira, apresentando em alguns instantes de tempo. Quando se compreende oque esta acontecendo ela pode ou não, combinar com outras emoções dependendo do evento ocorrido, se combinar com outa emoção, tende a ser positiva ou negativa.

**2.4.2 *Alegria***

A alegria é uma emoção positiva que tem como finalidade a sensação de bem-estar subjetivo e satisfação com a vida. Esta emoção básica conduz a sentimentos positivos e potência a inibição de pensamentos negativos, pode ser utilizados para descrever o mesmo tipo de fenômeno como felicidade.

**2.4.3 *Tristeza***

A tristeza é uma emoção humano que expressa desânimo ou frustração em relação a alguém ou algo. Provém do diagnóstico de uma doença grave e é equivalente a uma grande perda, como um objeto fonte de gratificação ou valorização pode causar reações físicas como depressão, choro e insônia. Uma pessoa desesperançosa apresenta distorções cognitivas que tende a acreditar que seus problemas são irreversíveis e fazem sua vida como se fossem ausentes de qualquer satisfação, predominando sentimento de culpa e tristeza. Esta emoção básica conduz a sentimentos negativos e potência a inibição de pensamentos positivos.

**2.4.4 *Medo***

O medo é uma emoção humano que expressa o alerta e o perigo. o medo e encontrado quando um aumento de intensidade bastante rápido na estimulação neuronal, por muita das vezes por uma sensação de perigo real ou imaginário. Esta emoção causa um forte transtorno fisiológico, o coração e a respiração ficam acelerados, os músculos se contraem, as mãos tremem. Estas sensações derivam do sistema nervoso simpático, adrenalina e noradrenalina que agem sobre o nosso corpo humano quando é apresentado o medo. Esta emoção básica conduz a sentimentos negativos e potência a inibição de pensamentos positivos.

***2.4.5 Desgosto ou nojo***

O desgosto ou nojo é uma emoção humano que expressa em ocasiões desagradáveis e aversivas, e na expressão facial, o lábio superior é repuxado para os lados e dá-se também um ligeiro franzimento do nariz, com objetivo de tapar as narinas contra um odor ofensivo ou cuspir um alimento venenoso. Ou seja, tendo repulsa a algo. Esta emoção básica conduz a sentimentos negativos e potência a inibição de pensamentos positivos.

***2.4.6 Raiva***

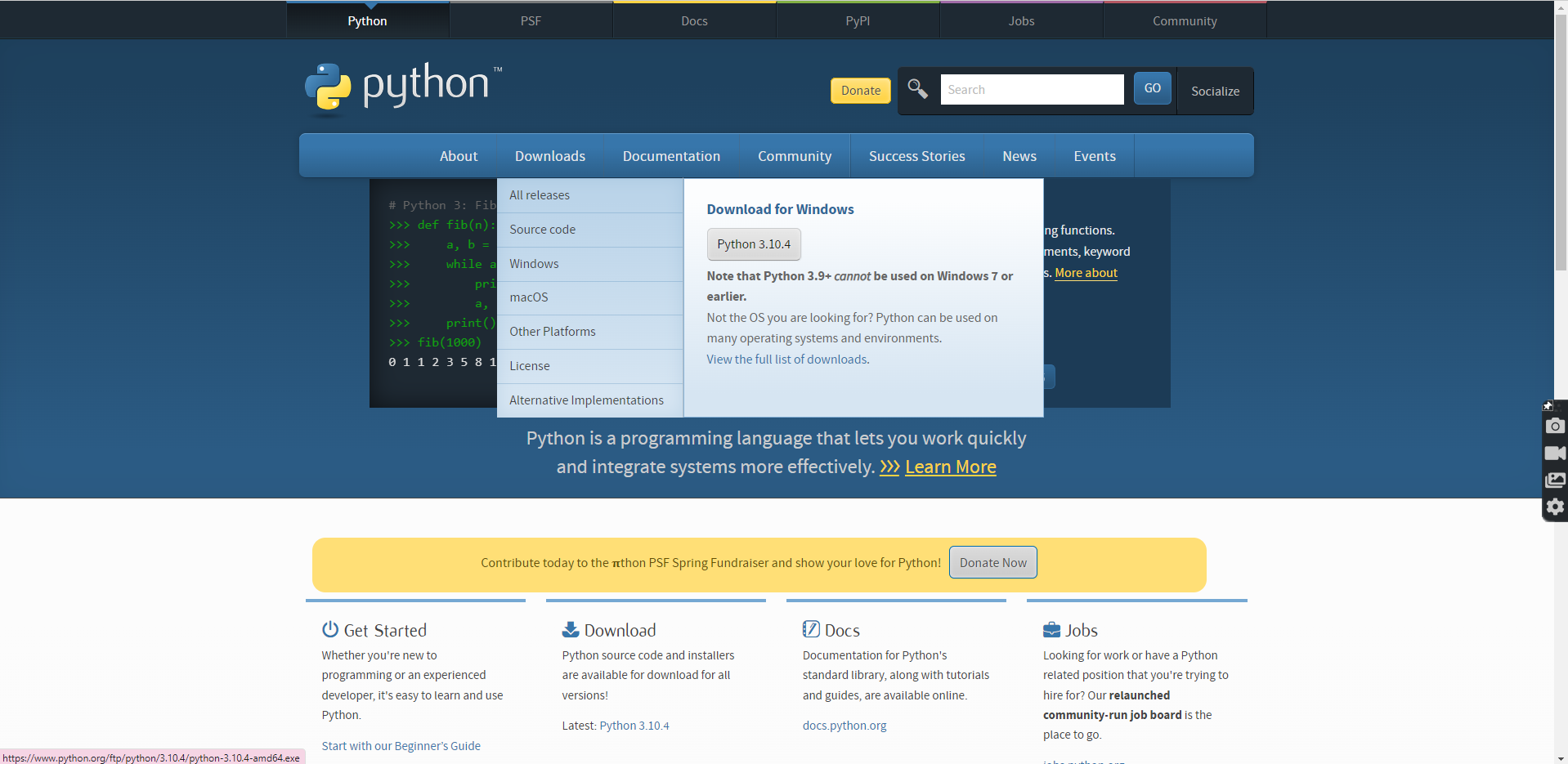
A raiva é uma emoção humana que expressa sensação de a revolta, a hostilidade, a irritabilidade, o ressentimento, a indignação, o ódio e a violência, sendo a única emoção que visa a ficar na defesa, pronta pra atacar com energias necessárias o seu alvo. Se acompanhar por reações físicas como o aumento da adrenalina e da noradrenalina no fluxo sanguíneo, aumento da pressão arterial e ritmo cardíaco, contração muscular e aceleração da respiração. Sendo a emoção mais perigosa, uma vez que conduz à tendência para prejudicar o alvo da que a levou a ter esta emoção. Esta emoção básica conduz a sentimentos negativos e potência a inibição de pensamentos positivos.

***3.* Requisitos da aplicação do projeto**

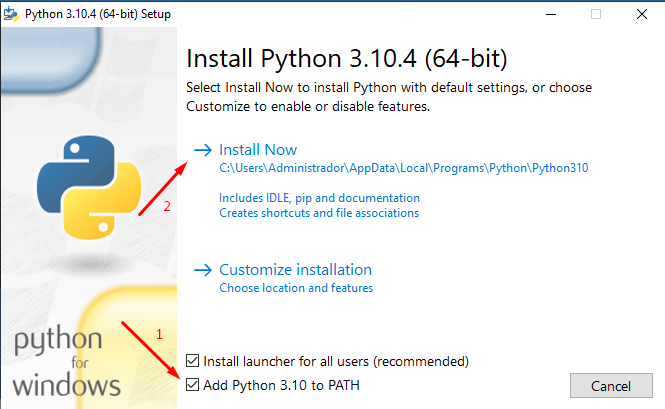
Os requisitos do estudo de caso de uso, será necessária a instalação do Python 3.5 ou superior, e download e importação da biblioteca NLTK, instalação de uma IDE e download dos pacotes NLTK. Neste estudo, foi utilizado o Sistema Operacional Windows 10.

**3.1.1 Primeiro passo a instalação do Python.**

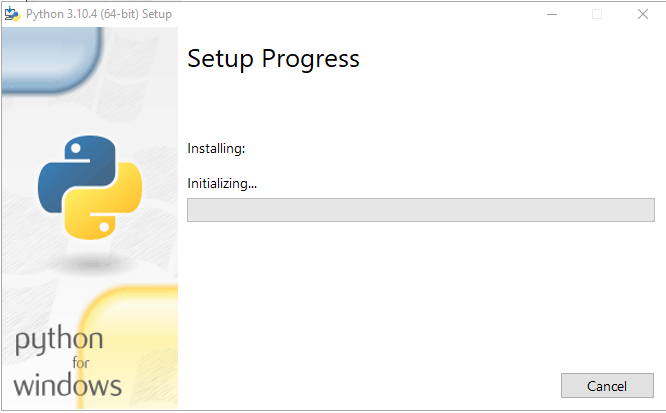
O primeiro passo e entrar no site: https://www.python.org/ na parte de downloads e baixar o programa instalador do python. Como na imagem a seguir.

Figura 4 – Site do Python, Fonte autor

Ao iniciar a instalação como na seta informada na imagem a baixo, marque a opção ADD PYTHON TO PATH para que as configurações do pip sejam entendiadas pelo prompt de comando, em seguida clique em Install Now.

Figura 5 – Foto do Instalador, Fonte autor

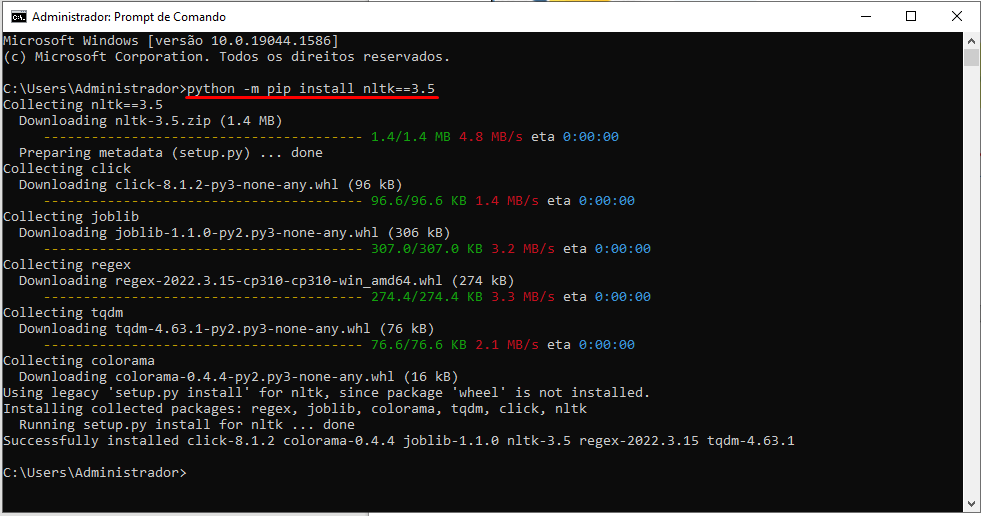
A instalação ira ser iniciada, aguardar o término da instalação.

Figura 6 – Foto do Instalador, Fonte autor

Como na imagem abaixo, clique na opção Disable path length limit, depois clique em close (fechar) como nas setas informadas na imagem abaixo, está opção não afetará nenhuma outra configuração do sistema. Ao ativar esta opção, isto resolverá possíveis problemas de comprimento de nome que podem surgir com projetos Python.

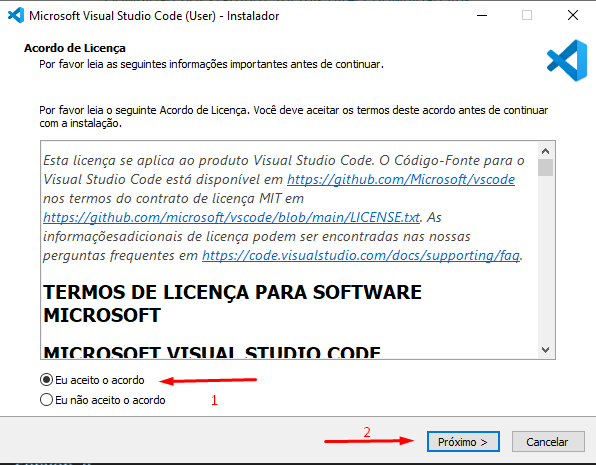
Figura 7 – Foto do Instalador, Fonte autor

Próximo passo abrir o promt de comando em modo de administrador e executar o comando: **“python -m pip install nltk==3.5”,** para instalar usando o comando pip o nltk versão 3.5, como na imagem abaixo.

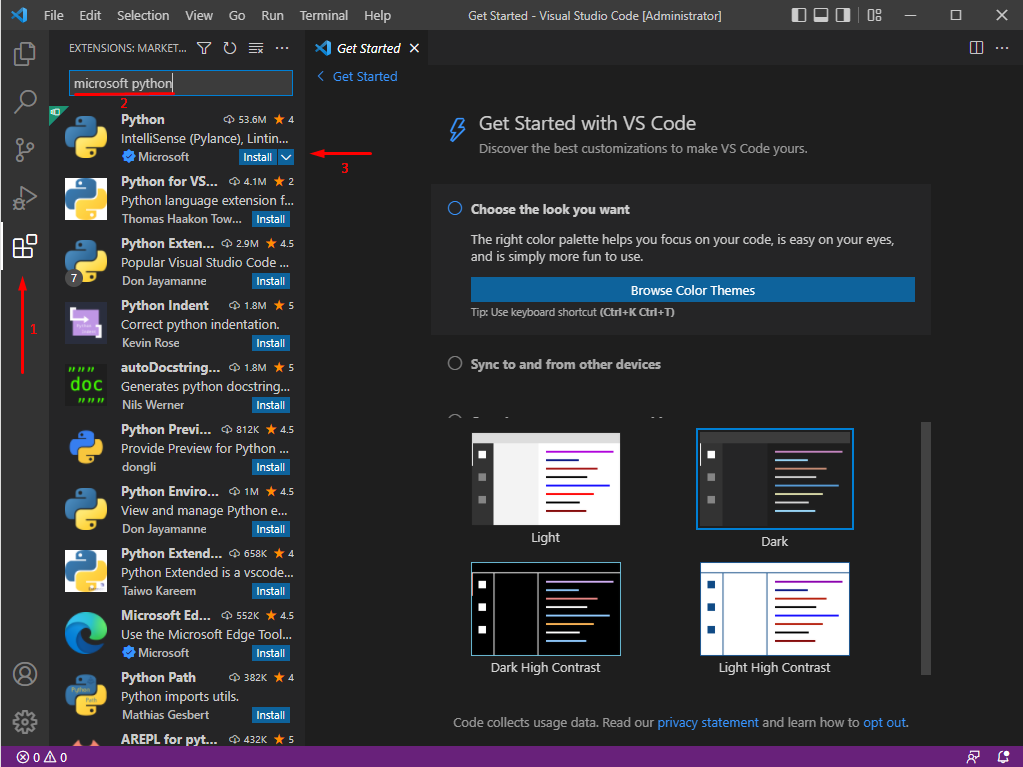
Figura 8 – Foto do Prompt de Comando, Fonte autor

**3.1.2 Segundo passo Instalação da IDE e dependência.**

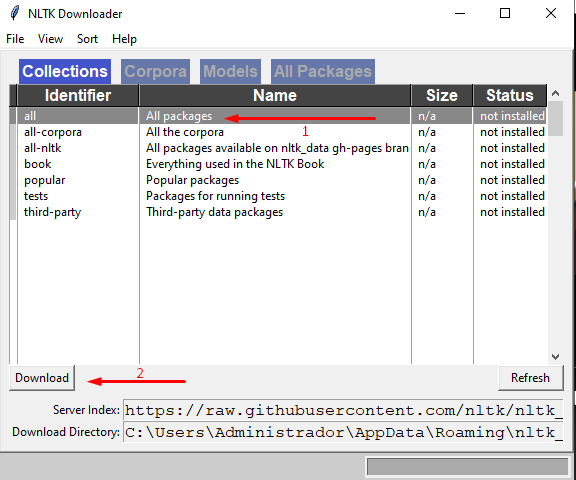
Próximo passo, será instalado a IDE do vscode e instalar a extensão Microsoft Python no site: https://code.visualstudio.com/docs/?dv=win apos o download faça a instalação. Seguindo o passo das imagens abaixo e avançar ao final concluir:

Figura 9 – Foto do Instalador VsCode – Fonte autor

Próximo passo, execute o VsCode, e clique em Extensions, escreva: “Microsoft Python” e clique em Install conforme a imagem abaixo:

Figura 10 – Foto do VsCode – Fonte autor

Concluído este processo, vá para a guia Terminal e para abrir o python, e digite Python em seguida importe o NLTK e faça o download da biblioteca NLTK. Com os comandos: “**import nltk**” e depois “**nltk.download()**”. Será aberta uma ferramenta de download do NLTK, para baixar a coleção de biblioteca NLTK: navegue pelos pacotes disponíveis, na guia collections no NLTK downloader mostra como os pacotes são agrupados em conjuntos, e você deve selecionar a linha chamada all packages para obter todos os dados necessários para os exercícios deste artigo. Em seguida clique em download, como na imagem a seguir:

Figura 11 – Foto do NLTK Download – Fonte autor

Depois que os dados são baixados para sua máquina, os requisitos foram alcançados para a realização do Estudo de Caso.

**3.2 Construção da Aplicação.**

Neste Estudo de Caso será feita a construção da aplicação para a realização dos estudos das emoções em maneiras linguísticas usando o Python com a biblioteca NLTK, com isso vamos inserir uma base de dados, disponibilizadas pelo autor Luís F. Dias e atualizada pelo autor Washington L. S. Menêzes. Com todas as frases já rotuladas para fazer que o algoritmo aprenda com os dois padrões abordados para frases de alegria, raiva, surpresa, medo, desgosto e tristeza. Que será dividida em duas variáveis sendo elas base de treinamento e base de teste, será necessária a divisão em duas bases para que o algoritmo verifique a assertividades em uma dessas bases para a esta finalidade. Para isso abra um novo arquivo e nomeie com Base.py, e cole a base de dados que se encontra anexada ao final desse artigo.

Próximo passo abra um novo arquivo e importe as dependências do projeto com os seguintes comandos:

import nltk

from nltk.tokenize import word\_tokenize

from nltk import pos\_tag

from nltk.metrics import ConfusionMatrix

from Base import BaseTreinamento

from Base import BaseTeste

from Teste import teste

#nltk.download()

Figura 12 – Importação de dependências - Fonte autor

Neste passo, será tratada as Stopwords, que são palavras de irrelevância para o algoritmo, como artigos, pronomes, verbos de estado como “ser” e “estar” e etc. São palavras que não são necessárias para o reconhecimento de padrões. o Stopwords pode ser utilizado em português com o seguinte comando:

stopwordsnltk = nltk.corpus.stopwords.words(‘portuguese’)

def removestopwords(texto):

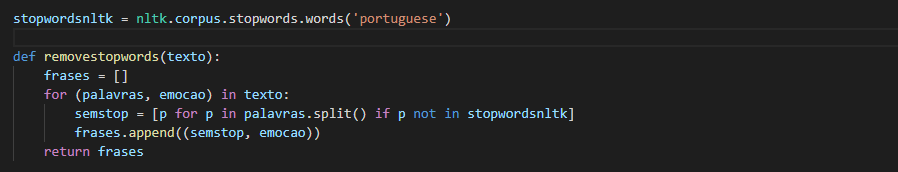
frases = []

for (palavras, emocao) in texto:

semstop = [p for p in palavras.split() if p not in stopwordsnltk]

frases.append((semstop, emocao))

return frases

Figura 13 – Implementação do Tokenize - Fonte autor

Neste passo, será realizo o Stemming, que é a técnica de remover as partes inúteis das palavras, deixando apenas seus radicais. Ex.: para Amor, Amado e Amável, o radical é Am. Para o algoritmo as 3 palavras seriam tratadas como uma só, fazendo com que o algoritmo seja mais dinâmico e menos carregado. Com o método abaixo nós retiramos as Stopwords da base de dados e ao mesmo tempo fazemos o processo de Stemming:

def aplicastemmer(texto):

stemmer = nltk.stem.RSLPStemmer()

frasesstemming = []

for (palavras, emocao) in texto:

comstemming = [str(stemmer.stem(p))

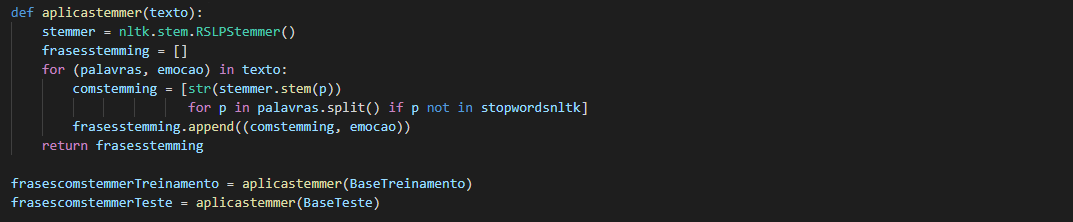
for p in palavras.split() if p not in stopwordsnltk]

frasesstemming.append((comstemming, emocao))

return frasesstemming

frasescomstemmerTreinamento = aplicastemmer(BaseTreinamento)

frasescomstemmerTeste = aplicastemmer(BaseTeste)

Figura 14 – Implementação do Stemming - Fonte autor

Próximo passo criaremos um método que busca todas as palavras da base de dados:

def buscaPalavras(frases):

todasPalavras = []

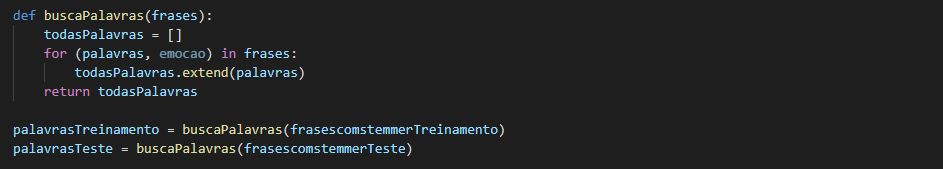
for (palavras, emocao) in frases:

todasPalavras.extend(palavras)

return todasPalavras

palavrasTreinamento = buscaPalavras(frasescomstemmerTreinamento)

palavrasTeste = buscaPalavras(frasescomstemmerTeste)

Figura 15 – Implementação do buscapalavra - Fonte autor

Próximo passo será criado o método abaixo, para ver as palavras mais frequentes na sua base de dados e efetuar mudanças necessárias:

def buscafrequencia(palavras):

palavras = nltk.FreqDist(palavras)

return palavras

frequenciatreinamento = buscafrequencia(palavrasTreinamento)

frequenciateste = buscafrequencia(palavrasTeste)

def buscaPalavrasUnicas(frequencia):

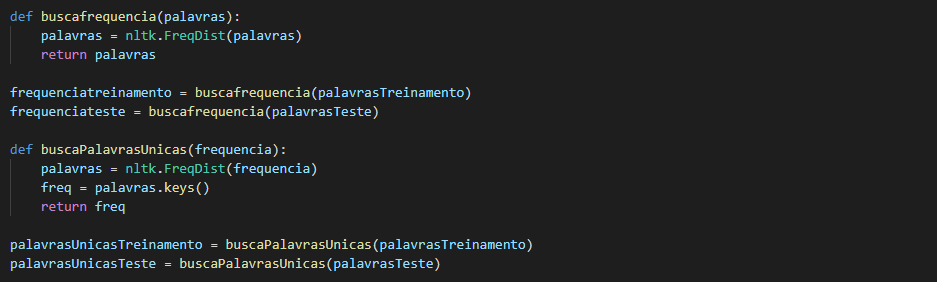
palavras = nltk.FreqDist(frequencia)

freq = palavras.keys()

return freq

palavrasUnicasTreinamento = buscaPalavrasUnicas(palavrasTreinamento)

palavrasUnicasTeste = buscaPalavrasUnicas(palavrasTeste)

Figura 16 – Implementação do buscafrequencia - Fonte autor

Próximo passo será criado um método para extrair palavras repetidas e logo em seguida utilize a biblioteca NLTK, para criar uma lista com todas as palavras tratadas até o momento:

def extratorPalavrasTreinamento(documento):

doc = set(documento)

caracteristicas = {}

for palavras in palavrasUnicasTreinamento:

caracteristicas['%s' % palavras] = (palavras in doc)

return caracteristicas

def extratorPalavrasTeste(documento):

doc = set(documento)

caracteristicas = {}

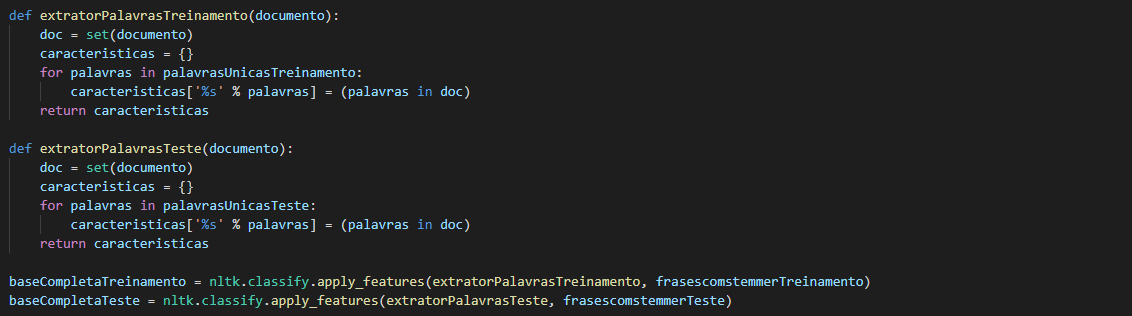
for palavras in palavrasUnicasTeste:

caracteristicas['%s' % palavras] = (palavras in doc)

return caracteristicas

baseCompletaTreinamento = nltk.classify.apply\_features(extratorPalavrasTreinamento, frasescomstemmerTreinamento)

baseCompletaTeste=nltk.classify.apply\_features(extratorPalavrasTeste,frasescomstemmerTeste)

Figura 17 – Implementação do ExtratorPalavrasTreinamento - Fonte autor

Próximo passo, utilize este comando da biblioteca NLTK, para construir a tabela de probabilidades e logo em seguida treinar seu algoritmo, também será criada a Tabela de Probabilidade dos recursos mais informativos do Classificador e testar a acurácia do algoritmo até o momento:

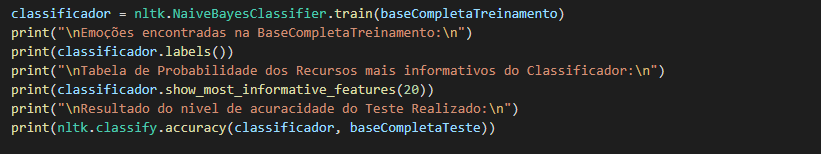
classificador = nltk.NaiveBayesClassifier.train(baseCompletaTreinamento)

print(classificador.labels())

print("\nTabela de Probabilidade dos Recursos mais informativos do Classificador:\n")

print(classificador.show\_most\_informative\_features(20))

print(nltk.classify.accuracy(classificador, basecompletateste))

Figura 18 – Implementação do Classificador - Fonte autor

O próximo passo será construir a matriz de confusão, que é uma forma intuitiva de saber como seu classificador está se comportando, já que só a acurácia não é suficiente para determinar a precisão de um algoritmo:

erros = []

for (frase, classe) in baseCompletaTeste:

resultado = classificador.classify(frase)

if resultado != classe:

erros.append((classe, resultado, frase))

esperado = []

previsto = []

for (frase, classe) in baseCompletaTeste:

resultado = classificador.classify(frase)

previsto.append(resultado)

esperado.append(classe)

matriz = ConfusionMatrix(esperado, previsto)

print("\nResultado do Matriz do Teste Realizado:\n")

print(matriz)

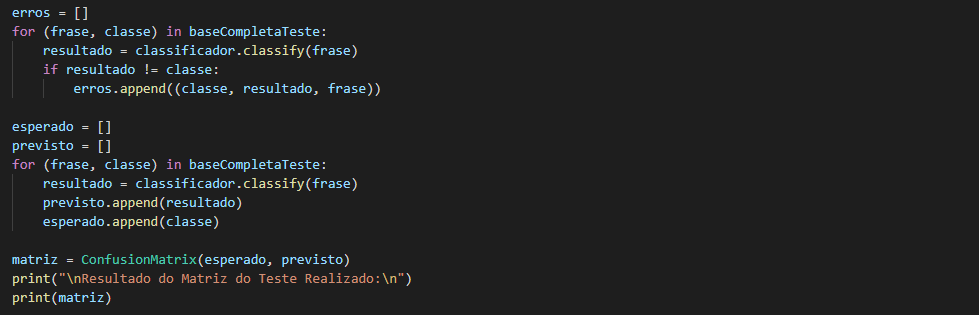


Figura 19 – Implementação da matriz de confusão - Fonte autor

Neste passo, será criado o metodo de stemming, no nas palavras de treinamento, com os comandos abaixo:

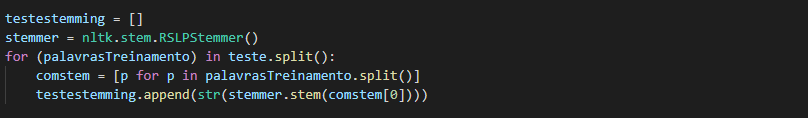
testestemming = []

stemmer = nltk.stem.RSLPStemmer()

for (palavrasTreinamento) in teste.split():

comstem = [p for p in palavrasTreinamento.split()]

testestemming.append(str(stemmer.stem(comstem[0])))

Figura 20 – Implementação do teste de Stemming - Fonte autor

Neste passo, será criado o método para imprimir os testes realizados, com os comandos abaixo:

novo = extratorPalavrasTeste(testestemming)

print("\nResultado do Stemming do Teste Realizado:\n")

print(novo)

distribuicao = classificador.prob\_classify(novo)

print("\nResultado do clasificador Naive Bayes do Teste Realizado:\n")

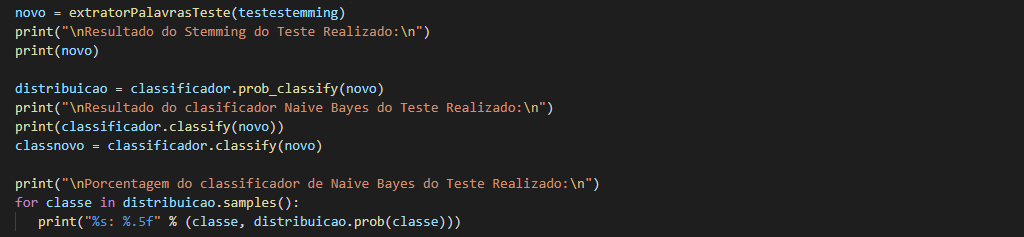
print(classificador.classify(novo))

classnovo = classificador.classify(novo)

print("\nPorcentagem do classificador de Naive Bayes do Teste Realizado:\n")

for classe in distribuicao.samples():

print("%s: %.5f" % (classe, distribuicao.prob(classe)))

Figura 21 – Implementação da impressão dos testes - Fonte autor

Próximo passo implementação do Tokenizador e a impressão do teste de sentenças do Tokenize, com os comandos abaixo:

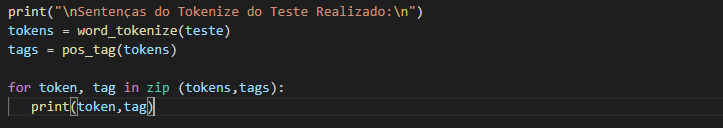
print("\nSentenças do Tokenize do Teste Realizado:\n")

tokens = word\_tokenize(teste)

tags = pos\_tag(tokens)

for token, tag in zip (tokens,tags):

print(token,tag)

Figura 22 – Implementação da impressão das sentenças dos Tokenize- Fonte autor

**3.2.1 Construção do teste.**

Próximo passo, implementaremos um novo arquivo, como o nome Teste.py, com a finalidade de digitar as frases a serem testadas com o nosso algoritmo, insira o comando abaixo, no entre aspas coloque a frase para ser classificada, exemplo: gostamos de ir a escola.

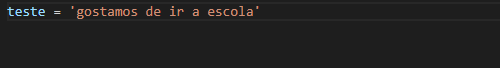
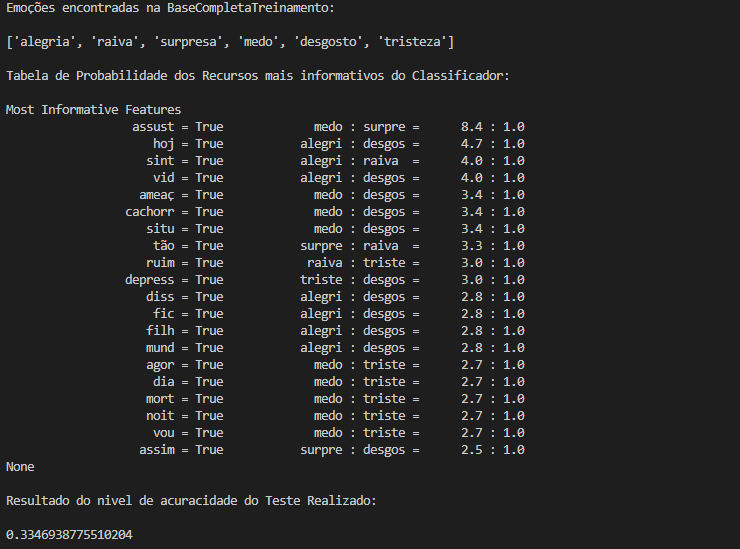
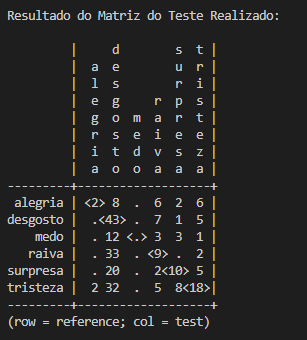
teste = 'gostamos de ir a escola'

Figura 23 – Implementação do Teste.py - Fonte autor

**4. Resultado dos testes**

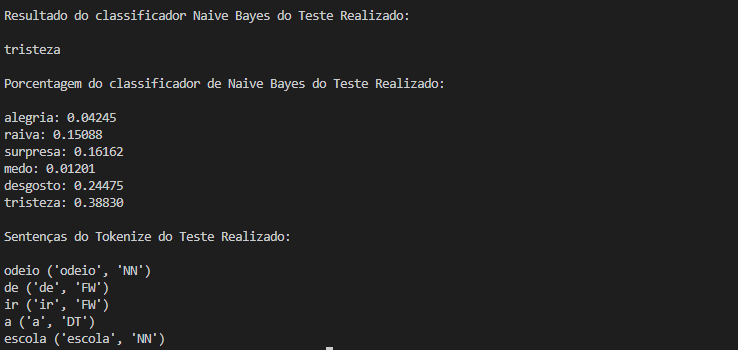
Foram realizados três testes com as frases, (odeio ir a escola), (gostamos de ir a escola), (não acredito que amanha vai fazer sol). e recebemos os resultados das emoções encontradas no banco de dados, a Tabela de Probabilidade dos Recursos mais informativos do Classificador, Resultado do nível de acuracidade do Teste Realizado, Resultado do Matriz, Resultado do Classificador Naive Bayes, Porcentagem do classificador de Naive Bayes, e as Sentenças do Tokenize do Teste Realizado.

Figura 24 – Resultado da acuracidade da BaseCompletaTreinamento - Fonte autor

Figura 25 – Resultado da Matriz de confusão da Base - Fonte autor

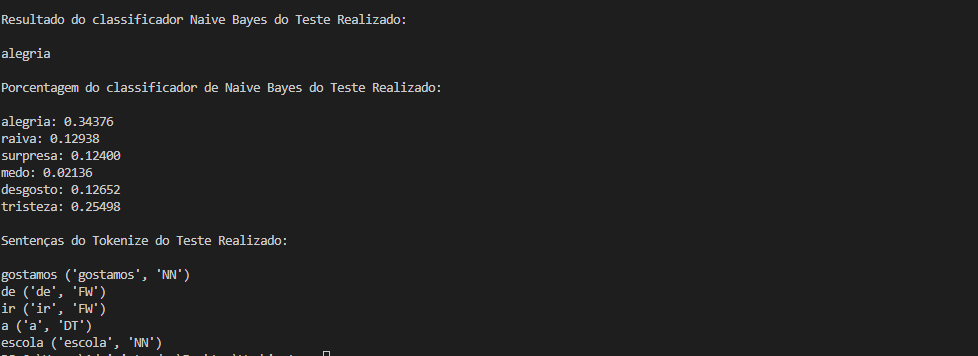
**4.1** ***Resultado “odeio ir a escola”***

Com o resultado do teste, obtivemos a resposta que a frase ***“odeio ir a escola”,*** significa tristeza, com 38% de assertividade, e que o Tokenizador possui substantivo massivo, frases estrangeiras, determinante.

Figura 26 – Resultado de “odeio ir a escola” - Fonte autor

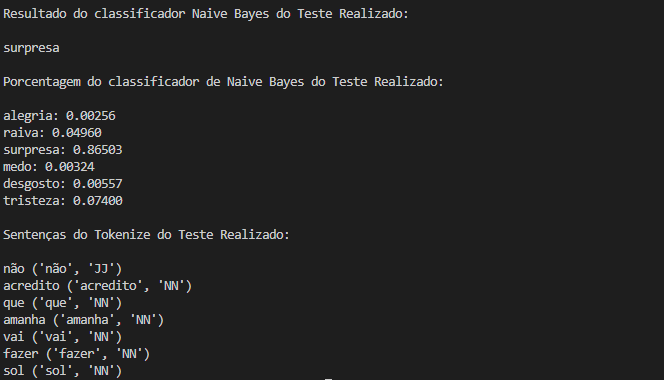
**4.2 *Resultado “gostamos de ir a escola”***

Com o resultado do teste, obtivemos a resposta que a frase ***“gostamos de ir a escola”,*** significa alegria, com 34% de assertividade, e que o Tokenizador possui substantivo massivo, frases estrangeiras, determinante.

Figura 27 – Resultado de “*gostamos de ir a escola”* - Fonte autor

**4.3 *Resultado “não acredito que amanha vai fazer sol”***

Com o resultado do teste, obtivemos a resposta que a frase ***“não acredito que amanha vai fazer sol”,*** significa Surpresa, com 86% de assertividade, e que o Tokenizador possui substantivo massivo e adjetivo.

Figura 28 – Resultado de *“não acredito que amanha vai fazer sol”*- Fonte autor

**5 CONSIDERAÇÕES FINAIS**

Podemos chegar a conclusão que ao inserir novos de dados na Base.py, e possível aumentar a precisão dos resultados, refinando a assertividade dos sentimentos abordados, assim como a porcentagem do classificador de Naive Bayes, que obtivemos sucesso nos testes realizados, em um dos três resultados obteve uma precisão de 86%.

Identificamos que este método e muito importante para tomadas de decisões, que pode-se descobrir o que o usuário está sentindo, tomando medidas para a solução de um problema ou opinião sobre algo abordado. Por exemplo em um chatbox com Inteligencia Artificial, e possível detectar a melhor abordagem a ser realizada com a resposta do sentimento.

**REFERÊNCIAS BIBLIOGRÁFICAS**

**Language Processing and Python.**

Disponivel em: https://www.nltk.org/book/ch01.html

**Natural Language Processing With Python's NLTK Package**

Disponivel em: https://realpython.com/nltk-nlp-python/

**NLP Tutorial Using Python NLTK (Simple Examples)**

Disponível em: https://likegeeks.com/nlp-tutorial-using-python-nltk/

**Examples from NLTK Book in Portuguese**

Disponível em: https://www.kaggle.com/code/leandrodoze/examples-from-nltk-book-in-portuguese/notebook

**Introdução ao Processamento de Linguagem Natural usando Python**

Disponível em: http://www.facom.ufu.br/~wendelmelo/terceiros/tutorial\_nltk.pdf

**Uma breve introdução ao teorema de Bayes para a ciência de dados**

Disponível em: https://rstudio-pubs-static.s3.amazonaws.com/529782\_fc03c44335b74250a9d5df1bb74fafbe.html

**Desenvolvimento de I.A. de Mineração de Emoção em Python com a biblioteca NLTK**

Disponível em: https://medium.com/@luisfdteixeira/desenvolvimento-de-i-a-de-minera%C3%A7%C3%A3o-em-python-com-a-biblioteca-nltk-4820a98d3db8

**As 6 únicas emoções presentes em todo o mundo**

Disponível em: https://www.psicologiamsn.com/2014/10/6-unicas-emocoes-presentes-em-todo-o-mundo-paul-ekman.html

**AGRADECIMENTOS:**

Agradeço a Deus por me guiar, me proteger, me instruir, pela minha saúde e por todas as bençãos em minha vida.

Agradeço à minha família, minha esposa Claudineia, e meus filhos maria luzia, maria alicia e lucas, por todo apoio prestado e também pela compreensão neste período.

Agradeço ao meu orientador, professor Sergio Assunção Monteiro, por todo apoio, motivação, paciência e dedicação ao longo do desenvolvimento deste trabalho, estando sempre disponível a compartilhar todo o seu vasto conhecimento.

**LISTA DO BANCO DE DADOS DA BASE.PY:**

BaseTreinamento = [

('eu gosto disso', 'alegria'),

('este trabalho e agradável','alegria'),

('gosto de ficar no seu aconchego','alegria'),

('fiz a adesão ao curso hoje','alegria'),

('eu sou admirada por muitos','alegria'),

('adoro como você e','alegria'),

('adoro seu cabelo macio','alegria'),

('adoro a cor dos seus olhos','alegria'),

('somo tão amáveis um com o outro','alegria'),

('sinto uma grande afeição por ele','alegria'),

('quero agradar meus filhos','alegria'),

('me sinto completamente amado','alegria'),

('eu amo você','alegria'),

('Viver feliz é viver de bem com a vida','alegria'),

('Sonhar, sorrir, viver e todo dia agradecer','alegria'),

('Melhor que rir, é rir com alguém ao seu lado','alegria'),

('Bonito é estar de bem com a vida','alegria'),

('O melhor lugar do mundo é onde você está em paz','alegria'),

('A vida é linda','alegria'),

('eu sou admirada por muitos','alegria'),

('me sinto completamente amado','alegria'),

('amar e maravilhoso','alegria'),

('estou me sentindo muito animado novamente','alegria'),

('eu estou muito bem hoje','alegria'),

('ele a feriu profundamente','raiva'),

('vou despejar minha cólera em você','raiva'),

('me sinto atormentado','raiva'),

('não me contrarie','raiva'),

('vou destruir tudo que foi construído','raiva'),

('não consigo terminar este trabalho, e muito frustrante','raiva'),

('me frustra a sua presença aqui','raiva'),

('esta comida me parece muito ruim','raiva'),

('você me destrói','raiva'),

('estamos separados','raiva'),

('estou odiando este vestido','raiva'),

('não pude comprar meu celular hoje','raiva'),

('ela e uma garota ruim','raiva'),

('estivemos em um show horroroso','raiva'),

('o ingresso estava muito caro','raiva'),

('se eu estragar tudo vai por água a baixo','raiva'),

('não possuo dinheiro algum','raiva'),

('sou muito pobre','raiva'),

('vai prejudicar a todos esta nova medida','raiva'),

('ficou ridículo','raiva'),

('este sapato está muito apertado','raiva'),

('a música e uma ofensa aos meus ouvidos','raiva'),

('não consigo terminar uma tarefa muito difícil','raiva'),

('reprovei em minha graduação','raiva'),

('estou muito chateado com tudo','raiva'),

('eu odeio em você','raiva'),

('e um desprazer conhecê-lo','raiva'),

('estou desperdiçando minhas ferias ','raiva'),

('e muito ruim este jogo','raiva'),

('vamos ter muito rancor pela frente','raiva'),

('não achei que seria tão terrível','raiva'),

('eu estou furioso com sua atitude','raiva'),

('minha raiva, estava me deixando furiosa, e descontrolada','raiva'),

(' eles merecem uma lição ','raiva'),

(' ainda estou muito bravo com você ','raiva'),

(' eu preciso surrar aquela chantagista ','raiva'),

(' olha o que você fez! derramou! ','raiva'),

(' você está pedindo pra apanhar! ','raiva'),

(' me deixa em paz! ','raiva'),

(' morra maldito, morra! ','raiva'),

(' você e mais irritante de perto ','raiva'),

(' e bom fechar o bico ','raiva'),

(' magicamente você me surpreendeu ','surpresa'),

(' e imenso esse globo ','surpresa'),

(' isso e tremendamente interessante ','surpresa'),

(' meu bilhete for sorteado, inacreditável! ','surpresa'),

(' um assalto a mão armada! ','surpresa'),

(' incrível, cabe em qualquer lugar! ','surpresa'),

(' você por aqui? ','surpresa'),

(' não dá pra acreditar no que ela me contou ','surpresa'),

(' os convidados já estão chegando! ','surpresa'),

(' puxa vida! nunca nos livramos de alguém tão depressa ','surpresa'),

(' micha carteira sumiu, eu estava com ela na mão ','surpresa'),

(' oh! um disco voador ','surpresa'),

(' amigos, que bela surpresa! ','surpresa'),

(' nunca pensei que veria isso e perto ','surpresa'),

(' nem acredito que comi tanto ','surpresa'),

(' não acredito que veio me ver ','surpresa'),

(' não acredito que e tão descarado ','surpresa'),

(' me surpreende sua falta de tato ','surpresa'),

(' o prédio onde eles moravam desabou! ','surpresa'),

(' inacreditável um bolo tão grande ','surpresa'),

(' e sério mesmo? não dá pra acreditar ','surpresa'),

(' como assim não vai ao nosso encontro? ','surpresa'),

(' como assim não tem ninguém em casa? ','surpresa'),

(' ué, mas para onde ele foi?! ','surpresa'),

(' por essa eu não esperava ','surpresa'),

(' nossa, olha só que mergulho ','surpresa'),

(' minha esposa está grávida! ','surpresa'),

(' meu dinheiro sumiu! ','surpresa'),

(' e verdade que os dois terminaram?!? ','surpresa'),

(' caramba, nem vi você chegar ','surpresa'),

(' nossa, como pode alguém cozinhar tão mal? ','surpresa'),

(' nossa que incrível ','surpresa'),

(' a fórmula sumiu! ','surpresa'),

(' eu nem acredito que já estou terminando o curso ','surpresa'),

(' não acredito que está aqui comigo novamente ','surpresa'),

(' está escondendo algo de nós! ','surpresa'),

(' como assim, ainda não terminou a tarefa ','surpresa'),

(' pensei que já estivesse pronta! ','surpresa'),

(' opa! quem apagou a luz? ','surpresa'),

(' caramba! aonde vai tão rápido? ','surpresa'),

(' estamos seguindo o caminho errado! ','surpresa'),

(' quatro reais o litro da gasolina! ','surpresa'),

(' me assustei ao vê-lo desse jeito! ','surpresa'),

(' minha mãe está grávida, acredita nisso? ','surpresa'),

(' parece mentira você ter crescido tanto ','surpresa'),

(' me surpreende sua imaginação ','surpresa'),

(' suas roupas são realmente lindas ','surpresa'),

('que abominável esse monstro!','medo'),

('vamos alarmar a todos sobre a situação','medo'),

('estou amedrontada','medo'),

('estou com muito medo da noite','medo'),

('ele está me ameaçando a dias','medo'),

('quanta angustia','medo'),

('estou angustiada','medo'),

('angustiadamente vou sair e casa','medo'),

('isso me deixa apavorada','medo'),

('você está me apavorando','medo'),

('estou desconfiada de você','medo'),

('não confio em você','medo'),

('até o cachorro está apavorado','medo'),

('estou assustado com as ações do meu colega','medo'),

('agora se sente humilhado, apavorado','medo'),

('assustou a população e provocou mortes','medo'),

('estou com dificuldades para respirar e muito assustado','medo'),

('os policiais se assustaram quando o carro capotou','medo'),

('o trabalhador e assombrado pelo temor do desemprego','medo'),

('eu estou assustado', 'medo'),

('não seja tão grosseiro','desgosto'),

('você fez uma manobra ilegal','desgosto'),

('sua indecente, não tem vergonha?','desgosto'),

('você e malvado com as crianças','desgosto'),

('que comentário maldoso','desgosto'),

('sem escrúpulos você manipula a tudo','desgosto'),

('sinto repulsa por você','desgosto'),

('e repulsivo a maneira como olha para as pessoas','desgosto'),

('estou indisposta','desgosto'),

('a indisposição me atacou hoje','desgosto'),

('acho que vou vomitar','desgosto'),

('tem muito vomito lá','desgosto'),

('que incomodo essa dor','desgosto'),

('não me incomode nunca mais','desgosto'),

('suas bobagens estão nos incomodando','desgosto'),

('que nojo olha toda essa sujeira','desgosto'),

('como isso está sujo','desgosto'),

('tenho náuseas só de lembrar','desgosto'),

('me sinto nauseada com o cheiro desta comida','desgosto'),

('você está obstruindo a passagem de ar','desgosto'),

('você está terrivelmente doente','desgosto'),

('olhe que feia esta roupa','desgosto'),

('que atitude deplorável','desgosto'),

('nossa como você e feio','desgosto'),

('muito mau tudo isso','desgosto'),

('estou desgostoso com você','desgosto'),

('você cortou o meu assunto','desgosto'),

('para que tanta chateação?','desgosto'),

('esse perfume e enjoativo','desgosto'),

('ser perigoso não nada bom','desgosto'),

('você e perigoso demais para minha filhas','desgosto'),

('que fedido este esgoto','desgosto'),

('que fedido você esta','desgosto'),

('que cachorro malcheiroso','desgosto'),

('hora que ultraje','desgosto'),

('e ultrajante da sua parte','desgosto'),

('situação desagradável essa','desgosto'),

('você só me da desgosto','desgosto'),

('tenho aversão a pessoas assim','desgosto'),

('antipatia e um mal da sociedade','desgosto'),

('que criatura abominável','desgosto'),

('e depressiva a maneira como você vê o mundo','desgosto'),

('me desagrada sua presença na festa','desgosto'),

('sinto asco dessa coisa','desgosto'),

('que hediondo!','desgosto'),

('vou golfar o cafe fora','desgosto'),

('hora que garota detestável!','desgosto'),

('estou nauseada','desgosto'),

('isso que você disse foi muito grave','desgosto'),

('não seja obsceno na frente das crianças','desgosto'),

('não seja rude com as visitas','desgosto'),

('esse assunto me da repulsa','desgosto'),

('que criança terrivelmente travessa','desgosto'),

('que criança mal educada','desgosto'),

('estou indisposta te dar o divorcio','desgosto'),

('tão patético, não tem nada mais rude para dizer?','desgosto'),

('por motivo torpe, com emprego de meio cruel e com impossibilidade de defesa para a vítima','desgosto'),

('a inveja e tão vil e vergonhosa que ninguém se atreve a confessá-la','desgosto'),

('o miserável receio de ser sentimental e o mais vil de todos os receios modernos','desgosto'),

('travesso gato quando fica com saudades do dono mija no sapato','desgosto'),

('isso e um ato detestável e covarde','desgosto'),

('revelam apenas o que e destrutivo e detestável para o povo','desgosto'),

('não sei como e a vida de um patife, mais a de um homem honesto e abominável','desgosto'),

('há coisas que temos que suportar para não acharmos a vida insuportável','desgosto'),

('as injurias do tempo e as injustiças do homem','desgosto'),

('odioso e desumano','desgosto'),

('você não publicará conteúdo odiento, pornográfico ou ameaçador','desgosto'),

('rancoroso e reprimido','desgosto'),

('não há animal mais degradante, estúpido, covarde, lamentável, egoísta, rancoroso e invejoso do que o homem','desgosto'),

('o virulento debate entre políticos','desgosto'),

('por favor não me abandone','tristeza'),

('não quero ficar sozinha','tristeza'),

('não me deixe sozinha','tristeza'),

('estou abatida','tristeza'),

('ele está todo abatido','tristeza'),

('tão triste suas palavras','tristeza'),

('seu amor não e mais meu ','tristeza'),

('estou aborrecida','tristeza'),

('isso vai me aborrecer ','tristeza'),

('estou com muita aflição','tristeza'),

('me aflige o modo como fala','tristeza'),

('estou em agonia com meu intimo','tristeza'),

('me sinto só','tristeza'),

('O seu amor só me trouxe lágrimas','tristeza'),

('A vida pode machucar muito mais que a morte','tristeza'),

('Está cada vez mais difícil esconder a dor','tristeza'),

('Chorar é a forma que o meu coração consegue se expressar','tristeza'),

('Decepcionado com a vida','tristeza'),

('São muitas decepções para um só coração','tristeza'),

('por favor não me abandone','tristeza'),

('não quero ficar sozinha','tristeza'),

('não me deixe sozinha','tristeza'),

('estou abatida','tristeza'),

('ele está todo abatido','tristeza'),

('tão triste suas palavras','tristeza'),

('seu amor não e mais meu','tristeza'),

('estou aborrecida','tristeza'),

('isso vai me aborrecer','tristeza'),

('estou com muita aflição','tristeza'),

('me aflige o modo como fala','tristeza'),

('estou em agonia com meu intimo','tristeza'),

('não quero fazer nada','tristeza'),

('me sinto ansiosa e tensa','tristeza'),

('não consigo parar de chorar','tristeza'),

('não consigo segurar as lagrimas','tristeza'),

('e muita dor perder um ente querido','tristeza'),

('estou realmente arrependida','tristeza'),

('acho que o carma volta, pois agora sou eu quem sofro','tristeza'),

('você não cumpriu suas promessas','tristeza'),

('me sinto amargurada','tristeza'),

('coitado esta tão triste','tristeza'),

('já e tarde de mais','tristeza'),

('nosso amor acabou','tristeza'),

('essa noite machuca só para mim','tristeza'),

('eu não estou mais no seu coração','tristeza'),

('você mudou comigo','tristeza'),

('quando eu penso em você realmente dói','tristeza'),

('como se fosse nada você vê minhas lagrimas','tristeza'),

('você disse cruelmente que não se arrependeu','tristeza'),

('eu nunca mais vou te ver','tristeza'),

('ela está com depressão','tristeza'),

('a depressão aflige as pessoas','tristeza'),

('estar depressivo e muito ruim','tristeza'),

('estou derrotada e deprimida depois deste dia','tristeza'),

('e comovente te ver dessa maneira','tristeza'),

]

BaseTeste =[

('não precisei pagar o ingresso','alegria'),

('se eu ajeitar tudo fica bem','alegria'),

('minha fortuna ultrapassa a sua','alegria'),

('sou muito afortunado','alegria'),

('e benéfico para todos esta nova medida','alegria'),

('ficou lindo','alegria'),

('achei esse sapato muito simpático','alegria'),

('estou ansiosa pela sua chegada','alegria'),

('congratulações pelo seu aniversário','alegria'),

('delicadamente ele a colocou para dormir','alegria'),

('a música e linda','alegria'),

('sem musica eu não vivo','alegria'),

('conclui uma tarefa muito difícil','alegria'),

('Simplesmente viva a vida','alegria'),

('Chorar é necessário, mas sorrir é essencial','alegria'),

('O seu sorriso pode mudar o dia de alguém','alegria'),

('Sorria, você foi feito para isso','alegria'),

('Todas as pessoas do mundo sorriem no mesmo idioma','alegria'),

('que belo dia para dirigir um carro novo','alegria'),

('o dia está muito bonito','alegria'),

('estou contente com o resultado do teste que fiz no dia de ontem','alegria'),

('o amor e lindo','alegria'),

('nossa amizade e amor vai durar para sempre', 'alegria'),

('eu amo você','alegria'),

(' com consegue ser tão bela? ','surpresa'),

(' essa e realmente uma casa deslumbrante ','surpresa'),

(' superou minhas expectativas ','surpresa'),

(' e admirável a maneira como se comporta ','surpresa'),

(' isso e realmente chocante ','surpresa'),

(' algumas noticias me surpreenderam no noticiário ','surpresa'),

(' surpreendente sua festa ','surpresa'),

(' estou tremendo de alegria ','surpresa'),

(' chocou grande parte do mundo ','surpresa'),

(' eu ficaria muito espantado com a sua vinda ','surpresa'),

(' ele e admirável ','surpresa'),

(' sua beleza me surpreendeu ','surpresa'),

(' seus olhos são surpreendentemente verdes ','surpresa'),

(' os políticos se surpreendem quando alguem acredita neles ','surpresa'),

(' estou perplexa com essas denuncias ','surpresa'),

(' fiquei perplexo com suas palavras ','surpresa'),

(' estou abismado com sua prosa ','surpresa'),

(' eu ficaria realmente abismado se me dissessem isso ','surpresa'),

(' o grupo foi surpreendido enquanto lavava o carro ','surpresa'),

(' estou boquiaberto com as imagens ','surpresa'),

(' estou boquiaberto com essas suas palavras ','surpresa'),

(' esse quadro e maravilhoso ','surpresa'),

(' este carro me deixou maravilhado ','surpresa'),

(' estou maravilhada ','surpresa'),

(' essa expectativa esta me matando ','surpresa'),

(' vou caminhar sempre na expectativa de encontrá-lo ','surpresa'),

(' você emudece minhas palavras ','surpresa'),

(' minhas palavras vão emudecer se não parar de me surpreender ','surpresa'),

(' a mulher e um efeito deslumbrante da natureza ','surpresa'),

(' estou deslumbrada com essas joias ','surpresa'),

(' isso e romântico e deslumbrante ','surpresa'),

(' isso pode ser surpreendentemente deslumbrante ','surpresa'),

(' trabalho deslumbrante ','surpresa'),

(' essas pessoas são esplêndida ','surpresa'),

(' e esplêndido como o céu se encontra no momento ','surpresa'),

(' e um carro fantástico ','surpresa'),

(' um edifício realmente fantástico ','surpresa'),

(' vou vetar o orçamento ao cliente ','raiva'),

(' meus pais não consentiram nosso casamento ','raiva'),

(' eu odiei este perfume ','raiva'),

(' seu descaso e frustrante ','raiva'),

(' me sinto completamente amarga ','raiva'),

(' desprezo muito o seu trabalho ','raiva'),

(' estamos descontentes por nossa família ','raiva'),

(' vou infernizar a sua empresa ','raiva'),

(' estou furioso com estes valores ','raiva'),

(' obrigaram o rapaz a sair ','raiva'),

(' como ele pode deixar de lado? ','raiva'),

(' são apenas injurias sobre mim ','raiva'),

(' estou enfurecido com a situação dessa empresa ','raiva'),

(' estou com o diabo no corpo ','raiva'),

(' isso foi diabólico ','raiva'),

(' tenho aversão à gente chata ','raiva'),

(' não vou perdoar sua traição ','raiva'),

(' esse dinheiro sujo e corrupto ','raiva'),

(' eles me crucificam o tempo todo ','raiva'),

(' eu vou enlouquecer com todo este barulho ','raiva'),

(' não aguento todo esse assedio ','raiva'),

(' cólera do dragão ','raiva'),

(' isso e ridículo! ','raiva'),

(' da próxima vez, vou inventar tudo sozinho ','raiva'),

(' seus tolos! deixaram ele escapar! ','raiva'),

(' jamais te perdoarei ','raiva'),

(' o que e isso? outra multa ','raiva'),

(' você passou dos limites! ','raiva'),

(' sente-se e cale a boca ','raiva'),

(' ingratos, vermes, vocês me pagam! ','raiva'),

(' saiam da dai, se não arranco vocês dai! ','raiva'),

(' você já me causou problemas suficientes ','raiva'),

(' isso foi a gota d'água ','raiva'),

(' o que você tem com isso? ','raiva'),

(' não vejo a hora de me livrar de você ','raiva'),

(' já entendi a jogada seus safados! ','raiva'),

(' você não merece piedade ','raiva'),

(' saia de perto de mim ','raiva'),

(' suma daqui, ou arranco seu couro! ','raiva'),

(' estou revoltado com essa situação ','raiva'),

(' seu idiota! ','raiva'),

(' não, eu não vou te emprestar dinheiro! ','raiva'),

(' você não passa de um cafajeste! vai embora ','raiva'),

(' pare de frescura e vá trabalhar ','raiva'),

('isso tudo e um erro','tristeza'),

('eu sou errada eu sou errante','tristeza'),

('tenho muito dó do cachorro','tristeza'),

('e dolorida a perda de um filho','tristeza'),

('essa tragédia vai nos abalar para sempre','tristeza'),

('perdi meus filhos','tristeza'),

('perdi meu curso','tristeza'),

('sou só uma chorona','tristeza'),

('você e um chorão','tristeza'),

('se arrependimento matasse','tristeza'),

('me sinto deslocado em sala de aula','tristeza'),

('foi uma passagem fúnebre','tristeza'),

('nossas condolências e tristeza a sua perda','tristeza'),

('Meu frágil coração cansou de te esperar','tristeza'),

('Minha ferida pode até sarar, mas sei que ela me deixará marcas','tristeza'),

('Nada nos deixa tão vulneráveis quanto a decepção','tristeza'),

('A pior solidão é estar com alguém e ainda assim se sentir solitário','tristeza'),

('e comovente ver o que os filhos do brasil passam','tristeza'),

('como me sinto culpada','tristeza'),

('estou abatida','tristeza'),

('a ansiedade tomou conta de mim','tristeza'),

('as pessoas não gostam do meu jeito','tristeza'),

('adeus passamos bons momentos juntos','tristeza'),

('sinto sua falta','tristeza'),

('ele não gostou da minha comida','tristeza'),

('estou sem dinheiro para a comida','tristeza'),

('queria que fosse o ultimo dia da minha vida','tristeza'),

('você está com vergonha de mim','tristeza'),

('ela não aceitou a minha proposta','tristeza'),

('era o meu ultimo centavo','tristeza'),

('reprovei de ano na faculdade','tristeza'),

('afinal você só sabe me desfazer','tristeza'),

('eu falhei em tudo nessa vida','tristeza'),

('eu fui muito humilhado','tristeza'),

('e uma história muito triste','tristeza'),

('ninguém acredita em mim','tristeza'),

('eu não sirvo para nada mesmo','tristeza'),

('droga, não faço nada direito','tristeza'),

('sofrimento em dobro na minha vida','tristeza'),

('fui demitida essa semana','tristeza'),

('as crianças sofrem ainda mais que os adultos','tristeza'),

('pra mim um dia e ruim, o outro e pior','tristeza'),

('de repente perdi o apetite','tristeza'),

('oh que dia infeliz','tristeza'),

('estamos afundados em contas','tristeza'),

('nem um milagre pode nos salvar','tristeza'),

('só me resta a esperança','tristeza'),

('pior que isso não pode ficar','tristeza'),

('meu salário e baixo','tristeza'),

('não passei no vestibular','tristeza'),

('ninguém se importa comigo','tristeza'),

('ninguém lembrou do meu aniversário','tristeza'),

('tenho tanto azar','tristeza'),

('o gosto da vingança e amargo','tristeza'),

('sou uma mulher amargurada depois de que você me deixou','tristeza'),

('estou desanimada com a vida','tristeza'),

('e um desanimo só coitadinha','tristeza'),

('a derrota e depressiva','tristeza'),

('discriminar e desumano','tristeza'),

('que desanimo','tristeza'),

('e uma desonra para o pais','tristeza'),

('a preocupação deveria nos levar a ação não a depressão','tristeza'),

('passamos ao desalento e a loucura','tristeza'),

('aquele que nunca viu a tristeza nunca reconhecerá a alegria','tristeza'),

('cuidado com a tristeza ela e um vicio','tristeza'),

('este lugar e mal assombrado','medo'),

('estou assombrado pela crise financeira','medo'),

('mesmo aterrorizado lembro de você','medo'),

('aterrorizado e suando frio','medo'),

('um grupo de elefantes selvagens tem aterrorizado vilas','medo'),

('me sinto intimidada pela sua presença','medo'),

('tenho medo de ser advertida novamente','medo'),

('estou correndo o risco de ser advertido','medo'),

('estou correndo riscos de saúde','medo'),

('os riscos são reais','medo'),

('podemos perder muito dinheiro com essa investida','medo'),

('socorro, fui intimado a depor','medo'),

('fui notificado e estou com medo de perde a guarda da minha filha','medo'),

('estou angustiada com meus filhos na rua','medo'),

('e abominável o que fazem com os animais','medo'),

('foi terrível o tigre quase o matou','medo'),

('me advertiram sobre isso','medo'),

('eu me assustei com você','medo'),

('eu estou apavorado','medo'),

('o mundo e feio como o pecado','desgosto'),

('a coisa mais difícil de esconder e aquilo que não existe','desgosto'),

('você errou feio aquele gol','desgosto'),

('nunca vou me casar sou muito feia','desgosto'),

('os golpes da adversidade são terrivelmente amargos','desgosto'),

('os homens ficam terrivelmente chatos','desgosto'),

('abominavelmente convencido','desgosto'),

('terrivelmente irritado','desgosto'),

('as instituições públicas estão terrivelmente decadentes','desgosto'),

('a população viveu em isolamento por muito tempo','desgosto'),

('estou terrivelmente preocupada','desgosto'),

('o nacionalismo e uma doença infantil','desgosto'),

('se me es antipático a minha negação esta pronta','desgosto'),

('muitos documentários sobre esse casal antipático','desgosto'),

('sua beleza não desfaça sua antipatia','desgosto'),

('esta e uma experiência desagradável','desgosto'),

('desagradável estrago nos banheiros','desgosto'),

('o mais irritante no amor e que se trata de um crime que precisa de um cúmplice','desgosto'),

('a situação nos causa grande incomodo','desgosto'),

('estou preocupado com o incomodo na garganta','desgosto'),

('simplesmente não quero amolação da policia','desgosto'),

('você e uma criaturinha muito impertinente','desgosto'),

('o peso e a dor da vida','desgosto'),

('me arrependo amargamente de minhas ações','desgosto'),

('o destino e cruel e os homens não são dignos de compaixão','desgosto'),

('o ódio conduz ao isolamento cruel e ao desespero','desgosto'),

('encerrou com o massacre mais repudiável e asqueroso que se conhece','desgosto'),

('de mal gosto e asqueroso','desgosto'),

('tudo e inserto neste mundo hediondo','desgosto'),

('o crime de corrupção e um crime hediondo','desgosto'),

('o rio está fétido e de cor escura','desgosto'),

('muito lixo no rio o deixa malcheiroso','desgosto'),

('existe uma laranja podre no grupo e já desconfiamos quem e','desgosto'),

('foi de repente estou machucado e me sentindo enjoado','desgosto'),

('eu fiquei enojado','desgosto'),

('daqui alguns meses vou embora deste pais que já estou nauseado','desgosto'),

('você e abominável','desgosto'),

('abomino a maneira como você age','desgosto'),

('estou adoentado','desgosto'),

('meu pai esta adoentado','desgosto'),

('estamos todos doentes','desgosto'),

('essa situação e muito amarga','desgosto'),

('disse adeus amargamente','desgosto'),

('tenho antipatia por aquela pessoa','desgosto'),

('como pode ser tão antipática!','desgosto'),

('que horrível seu asqueroso','desgosto'),

('tenho aversão agente como você','desgosto'),

('isso tudo e só chateação','desgosto'),

('estou muito chateada com suas mentiras','desgosto'),

('tão desagradável','desgosto'),

('isso me desagrada completamente','desgosto'),

('te desagrada isso','desgosto'),

('estou com enjoos terríveis','desgosto'),

('todos estão enfermos','desgosto'),

('foi uma enfermidade terrível','desgosto'),

('isso e muito grave','desgosto'),

]

LISTA DE STOPWORKS:

| **Lista de Stopworks** | |
| --- | --- |
| **inglês** | **Português** |
| a | uma |
| about | sobre |
| above | acima |
| across | através |
| after | depois de |
| again | novamente |
| against | contra |
| all | todos |
| almost | quase |
| alone | sozinho |
| along | ao longo |
| already | já |
| also | além disso |
| although | apesar |
| always | sempre |
| among | entre |
| an | a |
| and | e |
| another | outro |
| any | qualquer |
| anybody | qualquer pessoa |
| anyone | qualquer um |
| anything | qualquer coisa |
| anywhere | qualquer lugar |
| are | está |
| area | área |
| areas | areas |
| around | por aí |
| as | como |
| ask | pergunte |
| asked | perguntei |
| asking | perguntando |
| asks | pergunta |
| at | a |
| away | longe |
| b | b |
| back | de volta |
| backed | apoiado |
| backing | apoio |
| backs | costas |
| be | estar |
| became | passou a ser |
| because | porque |
| become | tornar-se |
| becomes | torna-se |
| been | fui |
| before | antes |
| began | começasse |
| behind | atrás |
| being | ser |
| beings | seres |
| best | melhor |
| better | melhor |
| between | entre |
| big | grande |
| both | ambos |
| but | mas |
| by | por |
| c | c |
| came | veio |
| can | posso |
| cannot | não podes |
| case | caso |
| cases | casos |
| certain | certo |
| certainly | certamente |
| clear | claro |
| clearly | claramente |
| come | venha |
| could | poderia |
| d | d |
| did | fez |
| differ | diferir |
| different | diferente |
| differently | diferentemente |
| do | faz |
| does | faz |
| done | feito |
| down | baixa |
| downed | abatido |
| downing | derrubada |
| downs | baixos |
| during | durante |
| e | e |
| each | cada |
| early | cedo |
| either | ou |
| end | fim |
| ended | terminou |
| ending | final |
| ends | termina |
| enough | o suficiente |
| even | até |
| evenly | uniformemente |
| ever | sempre |
| every | cada |
| everybody | todo o mundo |
| everyone | todos |
| everything | tudo |
| everywhere | em toda parte |
| f | f |
| face | face |
| faces | rostos |
| fact | facto |
| facts | factos |
| far | longe |
| felt | sentido |
| few | poucos |
| find | encontrar |
| finds | encontra |
| first | primeiro |
| for | para |
| four | quatro |
| from | de |
| full | cheio |
| fully | totalmente |
| further | mais distante |
| furthered | promovido |
| furthering | promovendo |
| furthers | promove |
| g | g |
| gave | deu |
| general | geral |
| generally | geralmente |
| get | pegue |
| gets | fica |
| give | dar |
| given | dado |
| gives | dá |
| go | vai |
| going | indo |
| good | boa |
| goods | mercadorias |
| got | obteve |
| great | ótimo |
| greater | maior |
| greatest | maior |
| group | grupo |
| grouped | agrupado |
| grouping | agrupamento |
| groups | grupos |
| h | h |
| had | teve |
| has | tem |
| have | ter |
| having | tendo |
| he | ele |
| her | dela |
| here | aqui |
| herself | ela própria |
| high | alto |
| higher | superior |
| highest | maior |
| him | ele |
| himself | ele mesmo |
| his | dele |
| how | como |
| however | contudo |
| i | eu |
| if | e se |
| important | importante |
| in | em |
| interest | interesse |
| interested | interessado |
| interesting | interessante |
| interests | interesses |
| into | para dentro |
| is | é |
| it | isto |
| its | está |
| itself | em si |
| j | j |
| just | somente |
| k | k |
| keep | manter |
| keeps | mantém |
| kind | tipo |
| knew | sabia |
| know | conhecer |
| known | conhecido |
| knows | sabe |
| l | eu |
| large | ampla |
| largely | largamente |
| last | último |
| later | mais tarde |
| latest | mais recentes |
| least | menos |
| less | menos |
| let | deixei |
| lets | vamos deixar |
| like | gostar |
| likely | provável |
| long | longo |
| longer | mais longo |
| longest | mais longo |
| m | m |
| made | fez |
| make | faço |
| making | fazer |
| man | homem |
| many | muitos |
| may | maio |
| me | mim |
| member | membro |
| members | membros |
| men | homens |
| might | poderia |
| more | mais |
| most | a maioria |
| mostly | na maioria das vezes |
| mr | sr |
| mrs | sra |
| much | muito de |
| must | devo |
| my | minha |
| myself | eu mesmo |
| n | n |
| necessary | necessário |
| need | precisar |
| needed | necessário |
| needing | precisando |
| needs | necessidades |
| never | nunca |
| new | novo |
| newer | mais recente |
| newest | o mais novo |
| next | próximo |
| no | não |
| nobody | ninguém |
| non | não |
| noone | ninguém |
| not | não |
| nothing | nada |
| now | agora |
| nowhere | lugar algum |
| number | número |
| numbers | números |
| o | o |
| of | do |
| off | fora |
| often | frequentemente |
| old | velho |
| older | mais velho |
| oldest | mais antigo |
| on | em |
| once | uma vez |
| one | um |
| only | só |
| open | abrir |
| opened | aberto |
| opening | abertura |
| opens | abre |
| or | ou |
| order | ordem |
| ordered | encomendado |
| ordering | encomenda |
| orders | encomendas |
| other | de outros |
| others | outras |
| our | nosso |
| out | fora |
| over | sobre |
| p | p |
| part | parte |
| parted | se separaram |
| parting | despedida |
| parts | partes |
| per | por |
| perhaps | possivelmente |
| place | lugar, colocar |
| places | locais |
| point | ponto |
| pointed | apontou |
| pointing | apontando |
| points | pontos |
| possible | possível |
| present | presente |
| presented | apresentado |
| presenting | apresentar |
| presents | presentes |
| problem | problema |
| problems | problemas |
| put | colocar |
| puts | coloca |
| q | q |
| quite | bastante |
| r | r |
| rather | em vez |
| really | realmente |
| right | certo |
| room | sala |
| rooms | quartos |
| s | s |
| said | disse |
| same | mesmo |
| saw | serra |
| say | dizer |
| says | diz |
| second | segundo |
| seconds | segundos |
| see | vejo |
| seem | parece |
| seemed | parecia |
| seeming | aparente |
| seems | parece |
| sees | vê |
| several | de várias |
| shall | deve |
| she | ela |
| should | devemos |
| show | exposição |
| showed | mostrou |
| showing | mostrando |
| shows | shows |
| side | lado |
| sides | lados |
| since | desde a |
| small | pequeno |
| smaller | menor |
| smallest | menor |
| so | assim |
| some | alguns |
| somebody | alguém |
| someone | alguém |
| something | alguma coisa |
| somewhere | algum lugar |
| state | estado |
| states | estados |
| still | ainda |
| such | tal |
| sure | certo |
| t | t |
| take | leva |
| taken | ocupado |
| than | do que |
| that | naquela |
| the | a |
| their | seus |
| them | eles |
| then | então |
| there | lá |
| therefore | assim sendo |
| these | estes |
| they | eles |
| thing | coisa |
| things | coisas |
| think | pensar |
| thinks | acha |
| this | isto |
| those | essa |
| though | apesar |
| thought | pensamento |
| thoughts | pensamentos |
| three | três |
| through | através |
| thus | portanto |
| to | para |
| today | hoje |
| together | juntos |
| too | também |
| took | tomou |
| toward | em direção a |
| turn | virar |
| turned | virou |
| turning | girando |
| turns | voltas |
| two | dois |
| u | você |
| under | debaixo |
| until | até |
| up | acima |
| upon | sobre |
| us | nos |
| use | usar |
| used | usava |
| uses | usa |
| v | v |
| very | muito |
| w | w |
| want | quer |
| wanted | procurado |
| wanting | querendo |
| wants | quer |
| was | estava |
| way | caminho |
| ways | maneiras |
| we | nós |
| well | bem |
| wells | poços |
| went | foi |
| were | estavam |
| what | que |
| when | quando |
| where | onde |
| whether | se |
| which | qual |
| while | enquanto |
| who | quem |
| whole | todo |
| whose | de quem |
| why | porque |
| will | vai |
| with | com |
| within | dentro |
| without | sem |
| work | trabalhos |
| worked | trabalhado |
| working | trabalhando |
| works | trabalho |
| would | seria |
| x | x |
| y | y |
| year | ano |
| years | anos |
| yet | ainda |
| you | você |
| young | jovem |
| younger | mais jovem |
| youngest | mais jovem |
| your | seu |
| yours | sua |
| z | z |
| share | compartilhar |
| linkthese | linkthese |
| am | sou |
| aren't | não são |
| below | abaixo |
| can't | não pode |
| couldn't | não podia |
| didn't | não fez |
| doesn't | não faz |
| doing | fazendo |
| don't | não faça |
| hadn't | não tinha |
| hasn't | não tem |
| haven't | não tem |
| he'd | ele tinha |
| he'll | inferno |
| he's | ele é |
| here's | aqui está |
| hers | dela |
| how's | como está |
| i'd | identidade |
| i'll | eu vou |
| i'm | eu estou |
| i've | eu tenho |
| isn't | não é |
| it's | está |
| let's | vamos |
| mustn't | não deve |
| nor | nem |
| ought | deveria |
| ours | nosso |
| ourselves | nós mesmos |
| own | próprio |
| shan't | shan't |
| she'd | ela |
| she'll | ela vai |
| she's | ela é |
| shouldn't | não deveria |
| that's | isso é |
| theirs | deles |
| themselves | si mesmos |
| there's | há |
| they'd | eles tinham |
| they'll | eles vão |
| they're | eles estão |
| they've | eles já |
| wasn't | não foi |
| we'd | qua |
| we'll | bem |
| we're | estamos |
| we've | nós temos |
| weren't | não estava |
| what's | o que é |
| when's | quando é |
| where's | onde está a |
| who's | quem é |
| whom | o qual |
| why's | por que |
| won't | não vai |
| wouldn't | não faria |
| you'd | você gostaria |
| you'll | você vai |
| you're | você é |
| you've | você tem |
| yourself | você mesmo |
| yourselves | vocês mesmos |