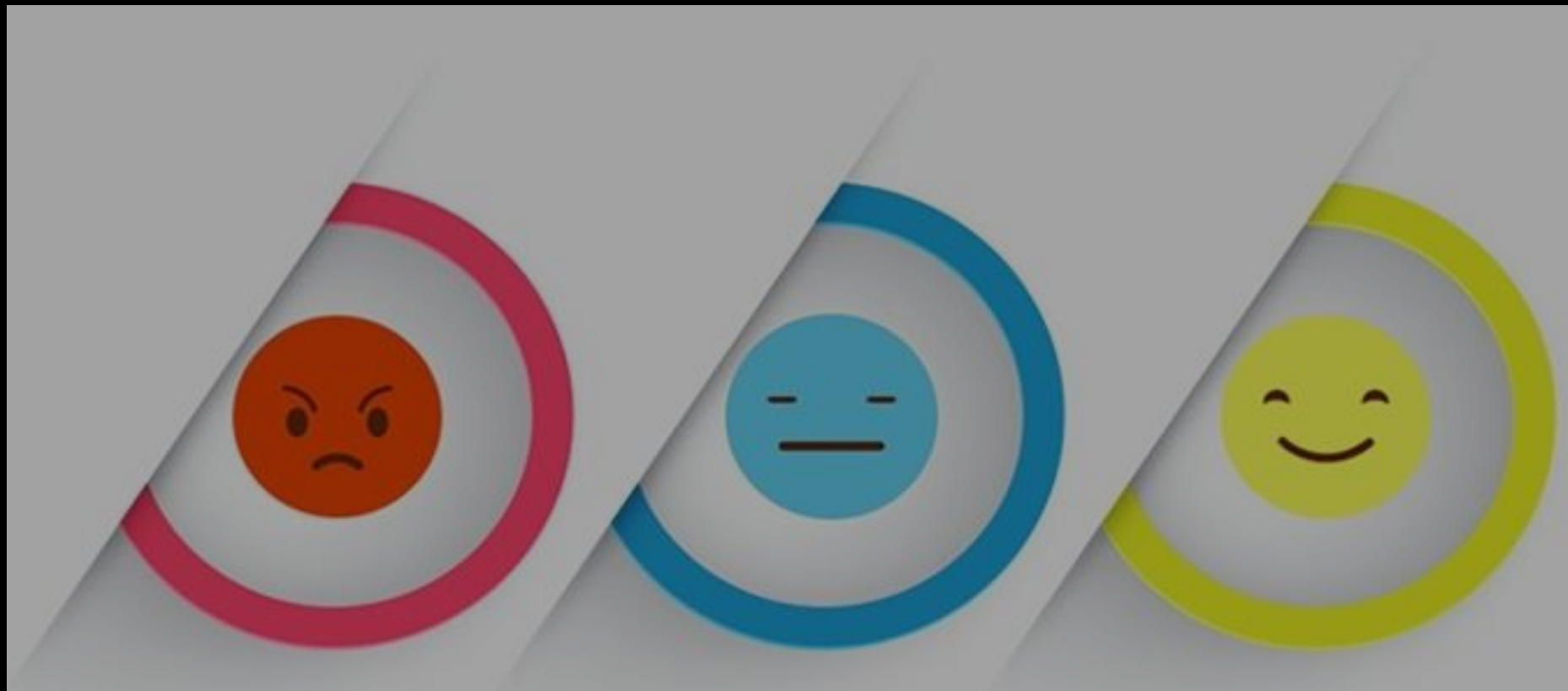




PUC Minas

Inteligência Artificial e Aprendizado de Máquina

Análise de Sentimentos

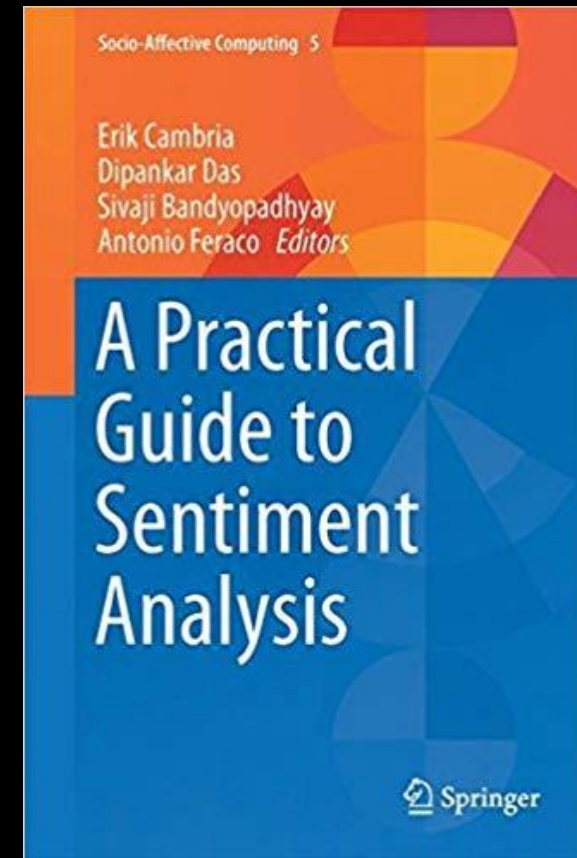
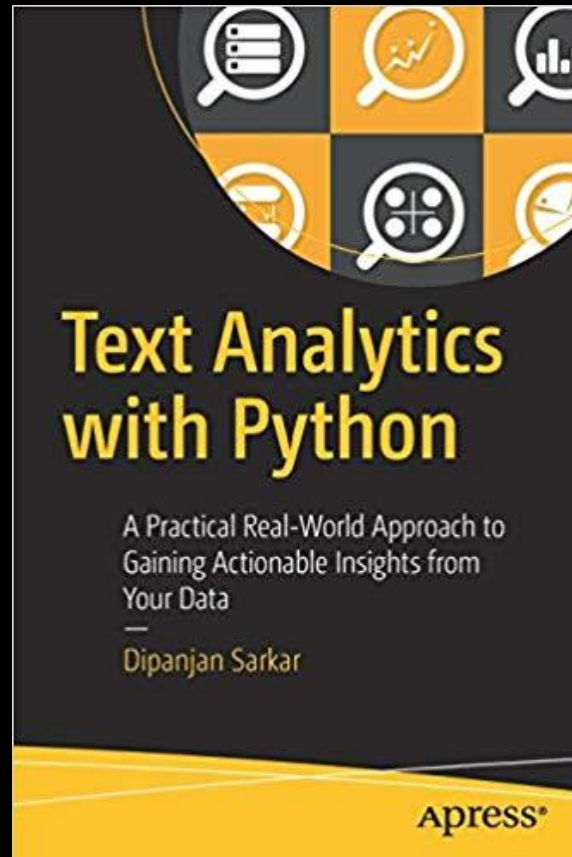


Prof. Otávio Augusto



PUC Minas

BIBLIOGRAFIA





PUC Minas

DISTRIBUIÇÃO DE PONTOS

- 60 pontos
 - Exercícios práticos / Aplicados a cada aula para fixação do conteúdo da aula
- 25 pontos
 - Trabalho prático / Construção de um modelo de Análise de Sentimentos com Gameificação
- 15 pontos
 - Seminários / Em grupo com temas relacionados à disciplina que não fazem parte da ementa



PUC Minas

ANÁLISE DE SENTIMENTOS

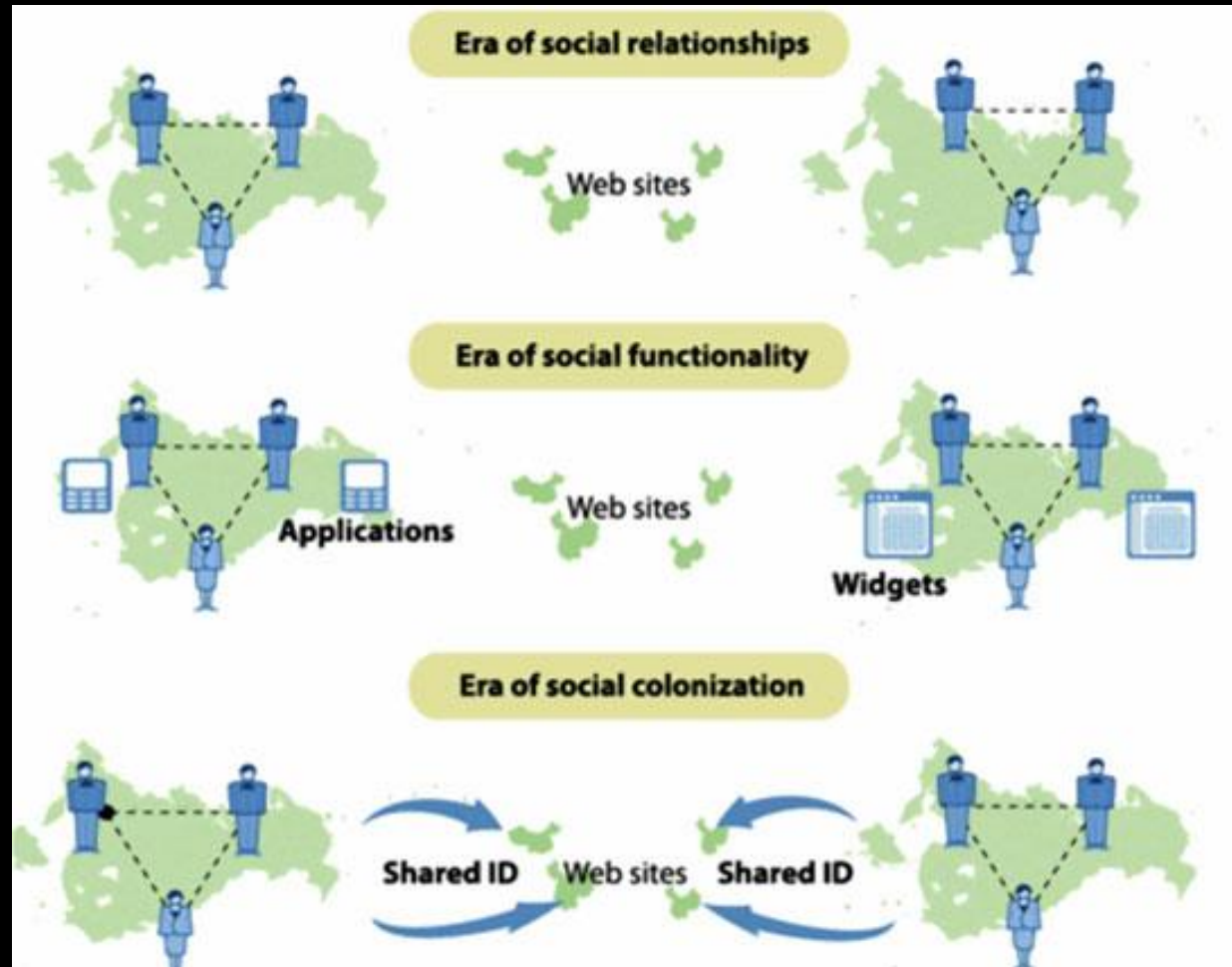
- MOTIVACÃO

- A passagem de uma Web de somente leitura para uma de escrita tornou os usuários mais entusiasmados em compartilhar suas emoções e opiniões por meio de redes sociais, comunidades on-line, blogs, wikis e outras mídias colaborativas on-line.
- Nos últimos anos, essa inteligência coletiva se espalhou para muitas áreas diferentes da Web, com foco particular em áreas relacionadas ao nosso cotidiano, como comércio, turismo, educação e saúde.



PUC Minas

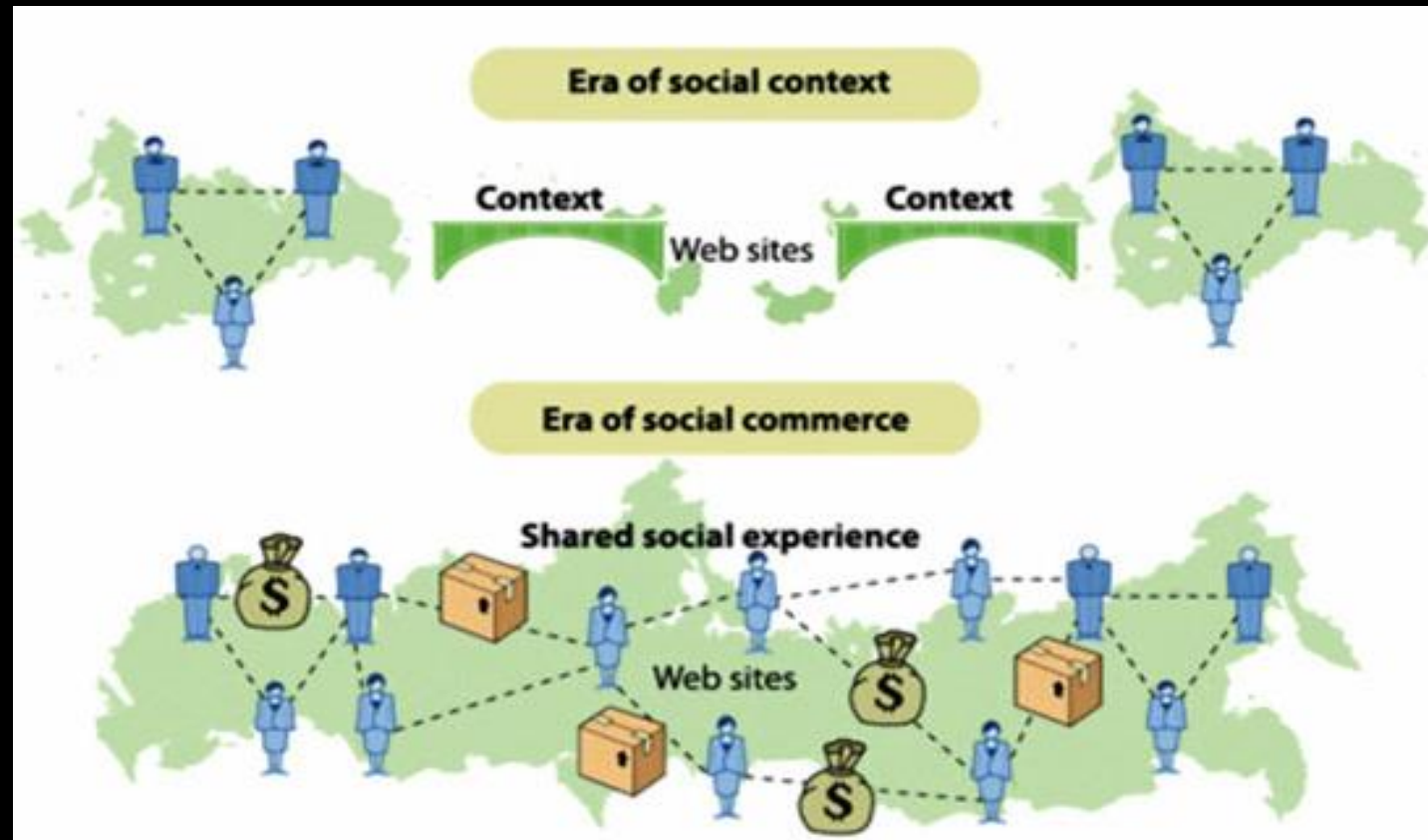
COMPUTAÇÃO AFETIVA E ANÁLISE DE SENTIMENTOS





PUC Minas

COMPUTAÇÃO AFETIVA E ANÁLISE DE SENTIMENTOS





PUC Minas

ANÁLISE DE SENTIMENTOS

- Entretenimento, jornalismo, mercado de ações, publicidade, soluções de negócio, reservar hotéis, política das redes sociais...

“Capturar e processar opiniões a fim de auxiliar um usuário ou uma empresa”
[Cambria et al., 2013]





PUC Minas

ANÁLISE DE SENTIMENTOS



Lewis Hamilton  @LewisHamilton · 4 de mai
I'm thankful for all your positivity and love 🙏 #TeamLH



POSITIVO

OU



NEGATIVO



PUC Minas

LINGUAGEM NATURAL

- Dados textuais são dados não estruturados
 - Geralmente pertencem a um idioma específico com sintaxe e semântica específicas
- Linguagem natural é uma linguagem criada e evoluída pelos seres humanos através do uso natural e da comunicação ao invés de ser construída e criada artificialmente, como uma linguagem de programação de computador.
- Linguagens naturais podem ser comunicadas de diferentes formas, incluindo fala, escrita ou até mesmo sinais.



PUC Minas

TRATAMENTOS DE TEXTOS PARA ANÁLISES NLP

- CORPUS

- Na tradução literal: “Corpo do texto”
 - É o plural em inglês de *text corpus*
 - São grandes e estruturados
 - Geralmente constituídos por corpos de textos escritos ou falados
 - Podem ser mono ou multilinguagem
- Objetivos
 - Aproveitá-los para fins linguísticos e estatísticos
 - Análise de dados e construções NLP.

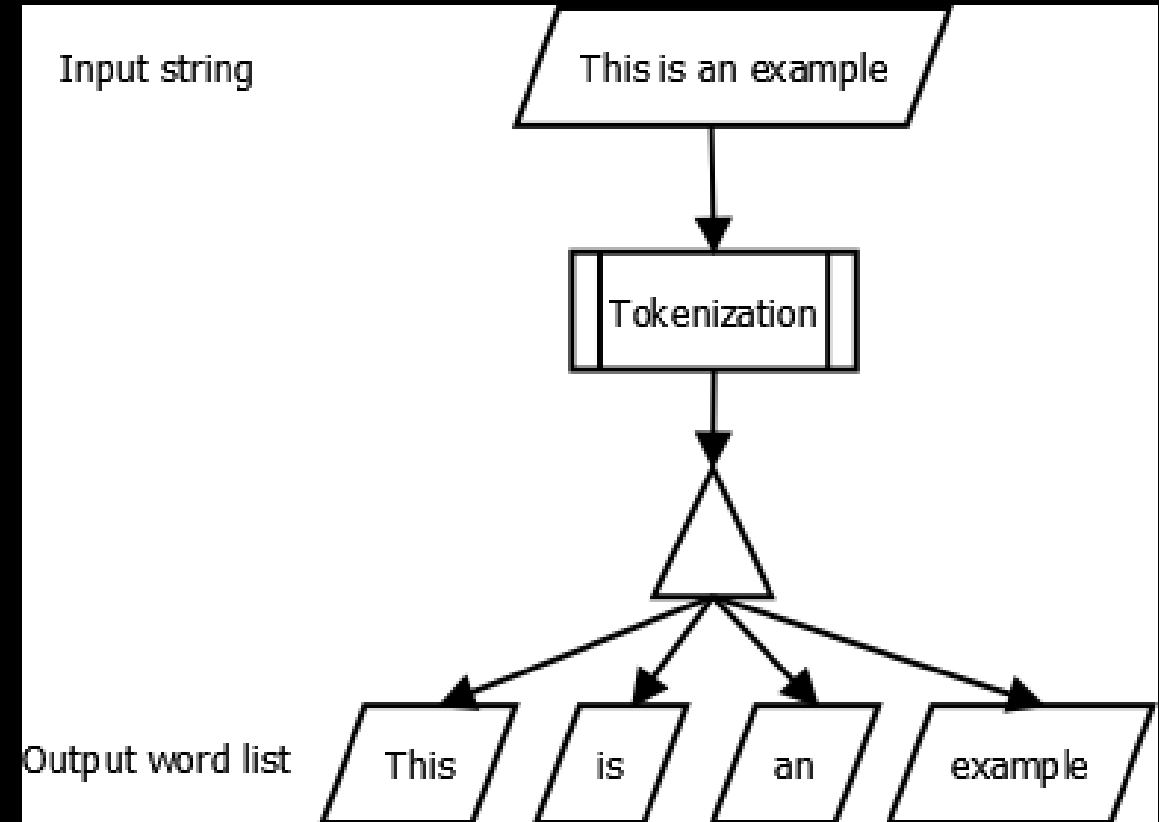


PUC Minas

TRATAMENTOS DE TEXTOS PARA ANÁLISES NLP

- TOKENS

- Termos individuais, como palavras. Um token é uma sequência de caracteres entre dois espaços, ou entre um espaço e sinais de pontuação. Um token pode ser também um número (seja ele real ou inteiro), ou ainda números separados por sinais de pontuação ou por caracteres, como ocorre em indicações de horas.





PUC Minas

TRATAMENTOS DE TEXTOS PARA ANÁLISES NLP

- NORMALIZAÇÃO

- “My Wife and I, and some friends, stayed here after this years grand national. We managed to book a double executive room for £80 and thought we were on to a winner - how wrong we were. After attending the national we arrived at the hotel to find out that there was no booking for us”



- “wife some friend stayed here after year grand national managed book double executive room thought were win how wrong were attend national we arrived hotel find out that there was no_book us”



PUC Minas

TRATAMENTOS DE TEXTOS PARA ANÁLISES NLP

- REMOÇÃO DE STOPWORDS
- Esse método consiste em remover palavras muito frequentes, tais como “a”, “de”, “o”, “da”, “que”, “e”, “do” entre outras, pois na maioria das vezes não são informações relevantes para a construção do modelo
- Remova *stopwords* somente quando não forem relevantes para a tarefa



PUC Minas

TRATAMENTOS DE TEXTOS PARA ANÁLISES NLP

- STEMIZAÇÃO
- *Stemming*
- Reduzir palavras flexionadas ou derivadas a sua base, ou seja, seu radical.
- “estudar, estudou, estudo e estudando” é reduzido ao termo “estud”
- “gato, gata, gatos e gatas” reduzem-se para “gat”



PUC Minas

TRATAMENTOS DE TEXTOS PARA ANÁLISES NLP

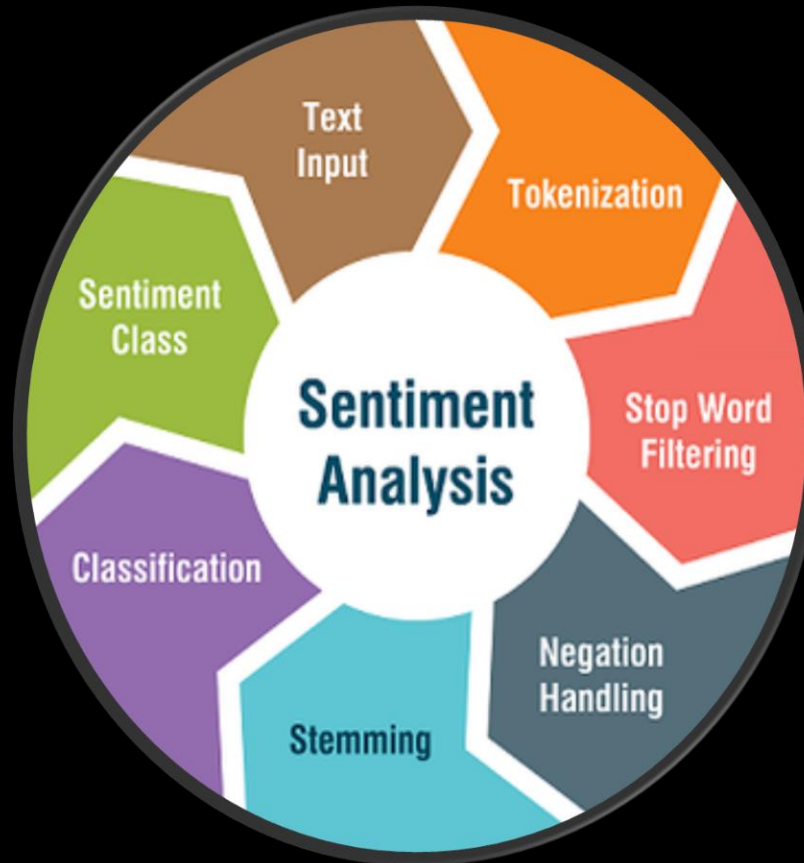
- LEMATIZAÇÃO

- *Lemming*
- Reduz a palavra ao seu lema, que é a forma no masculino e singular
- “gato, gata, gatos e gatas” são formas do lema “gato”
- “tiver, tenho, tinha e tem” são formas do lema “ter”



PUC Minas

ANÁLISE DE SENTIMENTOS





PUC Minas

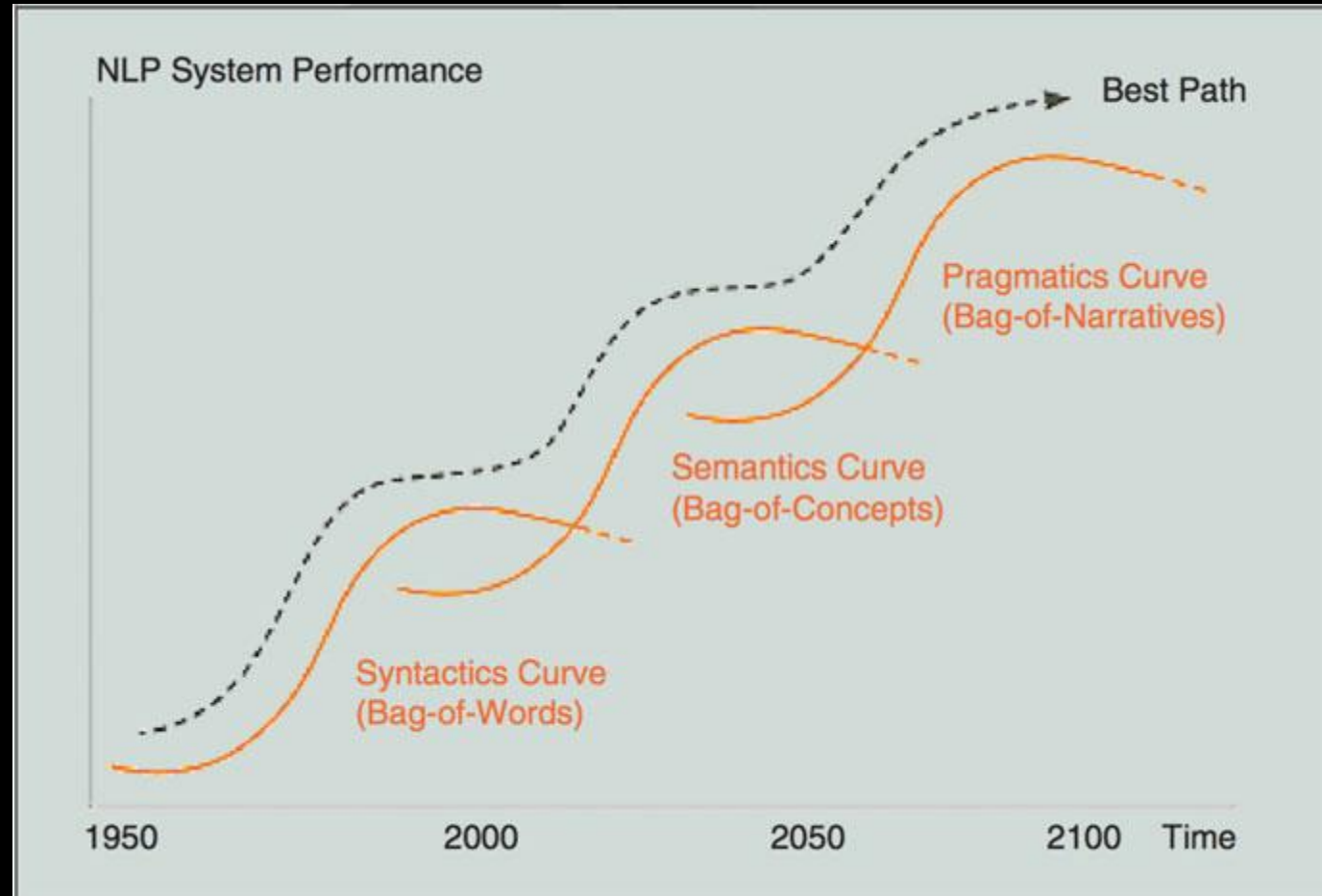
ANÁLISE DE SENTIMENTOS

- As tarefas básicas da computação afetiva e análise de sentimentos são reconhecimento de emoções e detecção de polaridade.
- Enquanto o primeiro se concentra na extração de um conjunto de rótulos de emoção, o segundo é geralmente uma tarefa de classificação binária com resultados como "positivo" vs. "negativo", "thumbs up" vs. "thumbs down" ou "like" vs. "dislike".



PUC Minas

ANÁLISE DE SENTIMENTOS





PUC Minas

PYTHON & ANÁLISE DE SENTIMENTOS

- Coisas amigáveis sobre o Python
 - Ele permite digitar diretamente no interpretador interativo
 - Execução pode ser feita com interfaces de terceiros
 - Ótima linguagem para análises estatísticas e ciência de dados
- Colaboratory
 - <https://colab.research.google.com>
 - É um ambiente do Jupyter Notebook gratuito que não requer configuração e é executado inteiramente na nuvem.



PUC Minas

PYTHON & ANÁLISE DE SENTIMENTOS

Welcome To Colaboratory

File Edit View Insert Run

CODE TEXT CELLS

contents Code snippets

g Colaboratory

started

ources

Learning Examples: Seedbank

N

SHARE

EDITING

AM

sk

ources, all for free from your

EXAMPLES RECENT GOOGLE DRIVE GITHUB UPLOAD

Filter notebooks

Title

- Overview of Colaboratory Features
- Markdown Guide
- Charts in Colaboratory
- External data: Drive, Sheets, and Cloud Storage
- Getting started with BigQuery

NEW PYTHON 3 NOTEBOOK CANCEL



PUC Minas

PYTHON & ANÁLISE DE SENTIMENTOS

 <https://www.nltk.org>

NLTK 3.4.3 documentation

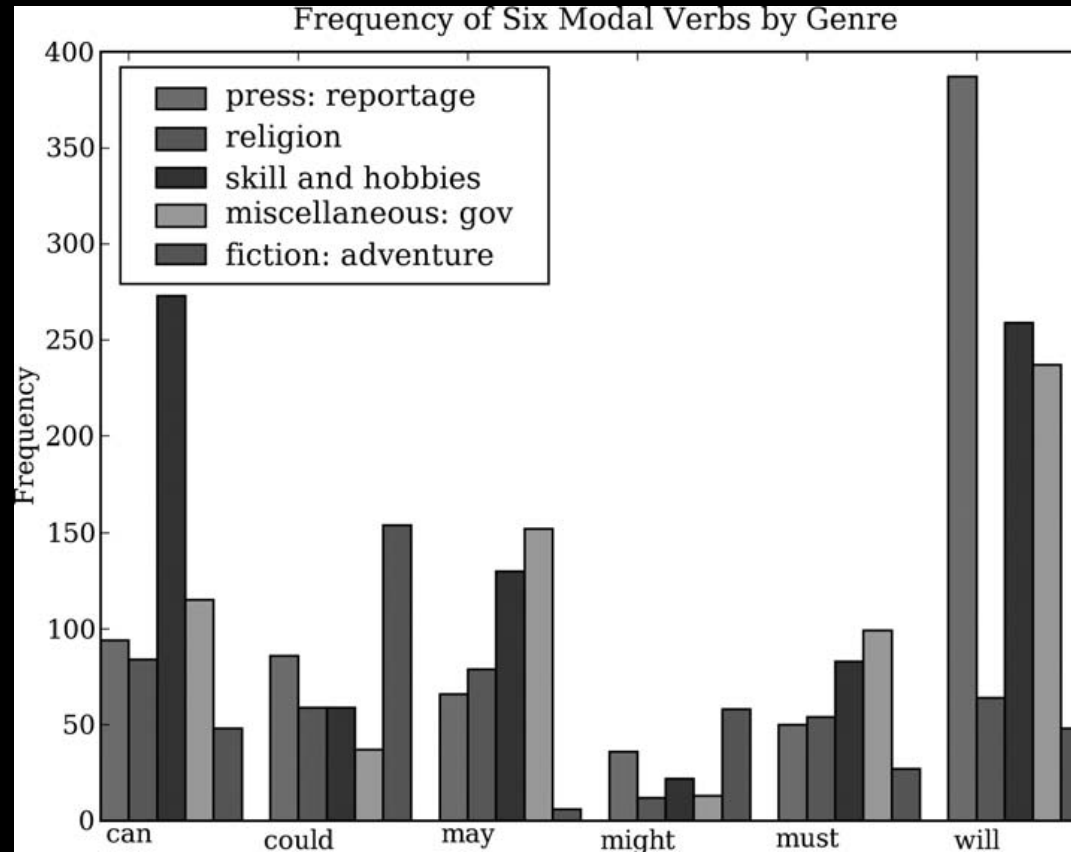
[NEXT](#) | [MODULES](#) | [INDEX](#)

Natural Language Toolkit



PUC Minas

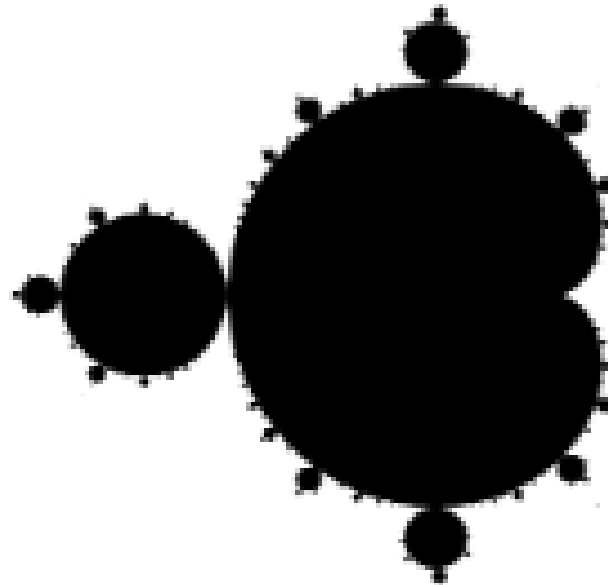
PYTHON & ANÁLISE DE SENTIMENTOS





PUC Minas

PYTHON & ANÁLISE DE SENTIMENTOS



TextBlob



PUC Minas

PYTHON & ANÁLISE DE SENTIMENTOS

Classification
Metrics in



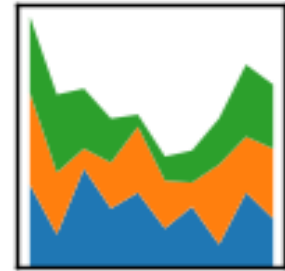
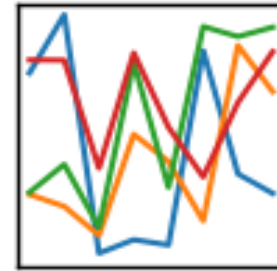
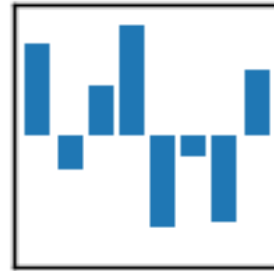


PUC Minas

PYTHON & ANÁLISE DE SENTIMENTOS

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$





PUC Minas

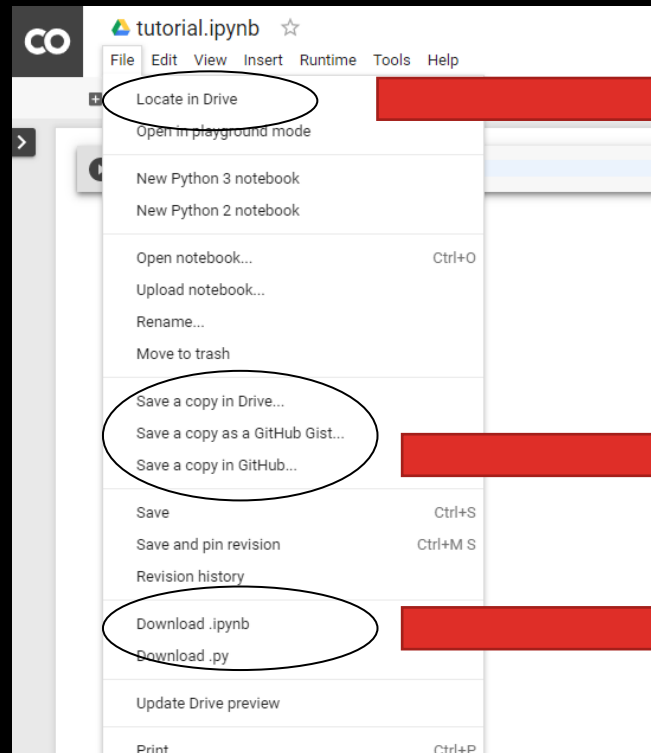
PYTHON & ANÁLISE DE SENTIMENTOS





PUC Minas

PYTHON & ANÁLISE DE SENTIMENTOS



Localizar o arquivo dentro do Google Drive

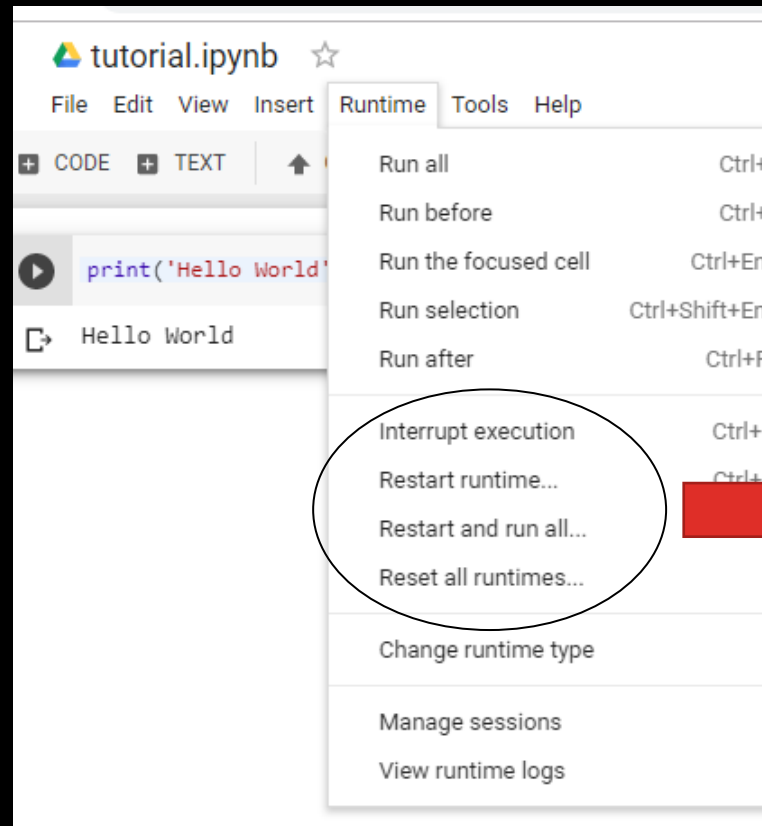
Salvar cópias do Colab no Drive ou no Github

Baixar o arquivo para a máquina (formato Jupyter ou Python)



PUC Minas

PYTHON & ANÁLISE DE SENTIMENTOS



- 1 – Interromper a execução de uma célula
- 2 – Reiniciar a execução
- 3 – Reiniciar e executar todas as células automaticamente
- 4 – Resetar as células executadas limpando a memória



PUC Minas

PYTHON & ANÁLISE DE SENTIMENTOS



Adiciona
uma
célula
para
código
em
Python

Adiciona
uma
célula
para
texto
livre

Ordena as
células
conforme
preferência
do usuário



PUC Minas

PYTHON & ANÁLISE DE SENTIMENTOS

Executa
o código
da célula

A screenshot of a Jupyter Notebook cell. The top part is a code editor with a play button icon on the left and the code `print('Hello World')`. The bottom part is an output area showing the text "Hello World" next to a copy icon.

Célula
para
escrita
do
código

Saída da
execução
do código

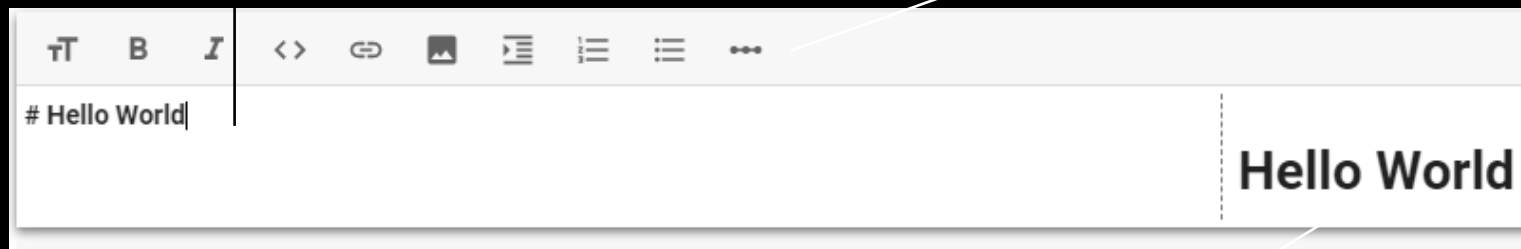


PUC Minas

PYTHON & ANÁLISE DE SENTIMENTOS

Célula para escrita do texto. O caractere # define o tamanho do texto, quantos mais houver, menor o texto será

Definição de fontes, marcadores, indentação ...



Prévia do que será impresso