

Tera

AULA 25

Fundamentos de Natural
Language Processing
(NLP)

Instrutor: André Barbosa (Bolinho)

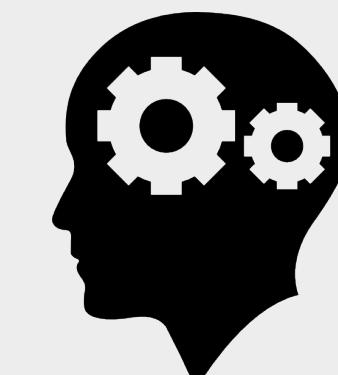
Background:

- Bacharel em Sistemas de Informação (EACH-USP)
 - Data Scientist – Elo7
-

Interesses:



Educação



Aprendizado
de Máquina



MOOCs



Information
Retrieval



Processamento de
Linguagem Natural



Sistemas de
recomendação

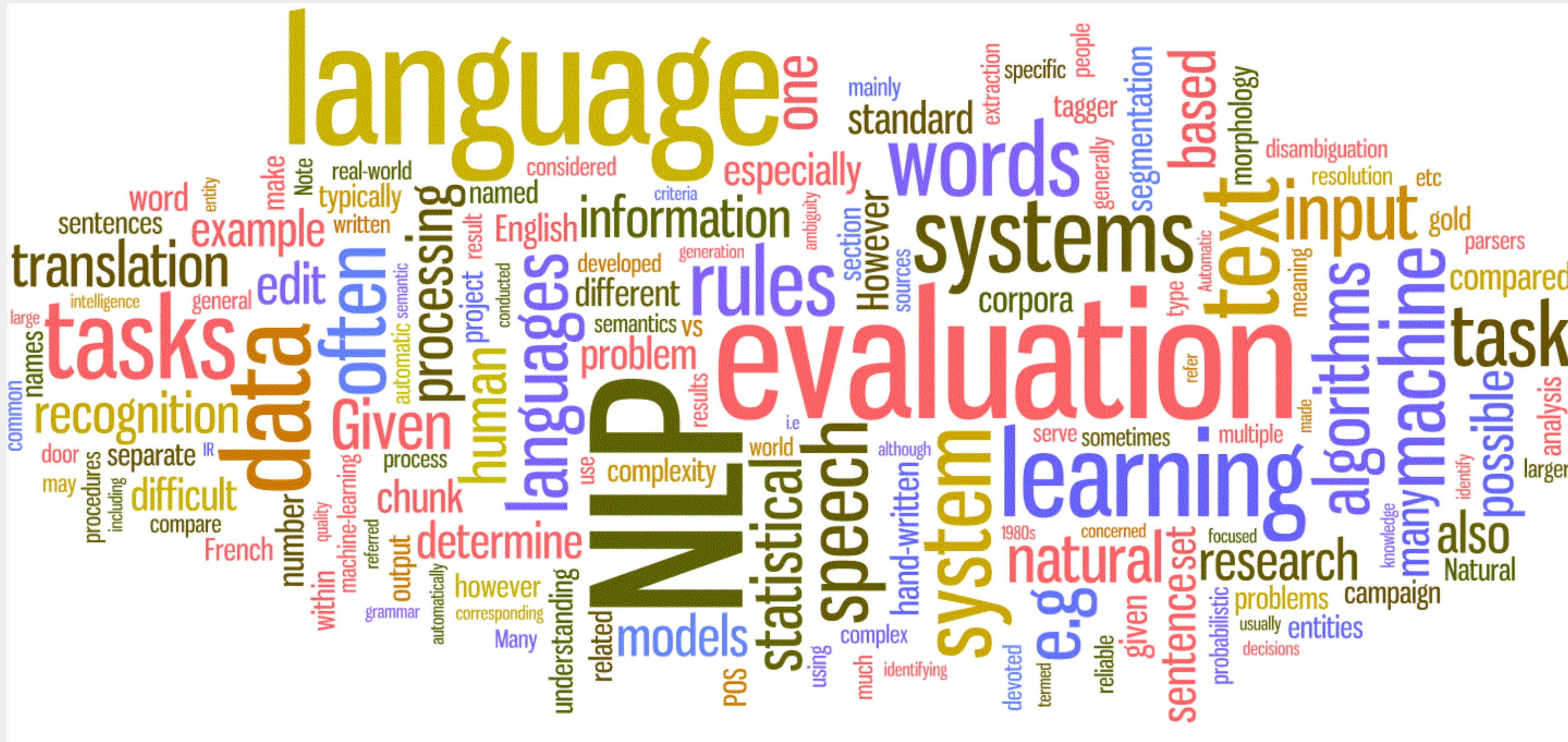
T

Planejamento:

1. O que é NLP para você?
2. Motivação
3. Jupyter Notebook !

T

O Que é NLP para você?



Motivação:

- Algumas aplicações



Dan Jurafsky

Question Answering: IBM's Watson

- Won Jeopardy on February 16, 2011!

WILLIAM WILKINSON'S
"AN ACCOUNT OF THE PRINCIPALITIES OF
WALLACHIA AND MOLDOVIA"
INSPIRED THIS AUTHOR'S
MOST FAMOUS NOVEL

Bram Stoker

Dracula

Motivação:

- Algumas aplicações

Dan Jurafsky



Information Extraction

Subject: **curriculum meeting**

Date: January 15, 2012

To: Dan Jurafsky

Hi Dan, we've now scheduled the curriculum meeting.

It will be in Gates 159 tomorrow from 10:00-11:30. ▾

-Chris

Create new Calendar entry

T

Motivação:

- # ➤ Algumas aplicações

Dan Jurafsky



Information Extraction & Sentiment Analysis



Attributes:

Attribute	Score (approx.)
zoom	85%
affordability	88%
size and weight	75%
flash	100%
ease of use	70%

Size and weight

- ✓ • nice and compact to carry!
- ✓ • since the camera is small and light, I won't need to carry around those heavy, bulky professional cameras either!
- ✗ • the camera feels flimsy, is plastic and very light in weight you have to be very delicate in the handling of this camera

Motivação:

- Algumas aplicações

Dan Jurafsky
Stanford University NLP Natural Language Processing

Machine Translation

- Fully automatic
- Helping human translators

Enter Source Text:
تعزز الرئيس اللبناني بول نهود لـ # حملة عنيفة في مجلس النواب الذي انعقد امس في جلسة تشريعية عادية تحررت
الى "محكمة" لـ # رئيس الجمهورية على موقف # من المحكمة الدولية و "العلامات" التي ادى بـ # لها
حول هذا الموضع .

Enter Translation:
This is only a matter of time.

Translation from Stanford's Phrasal:
This is only a matter of time.

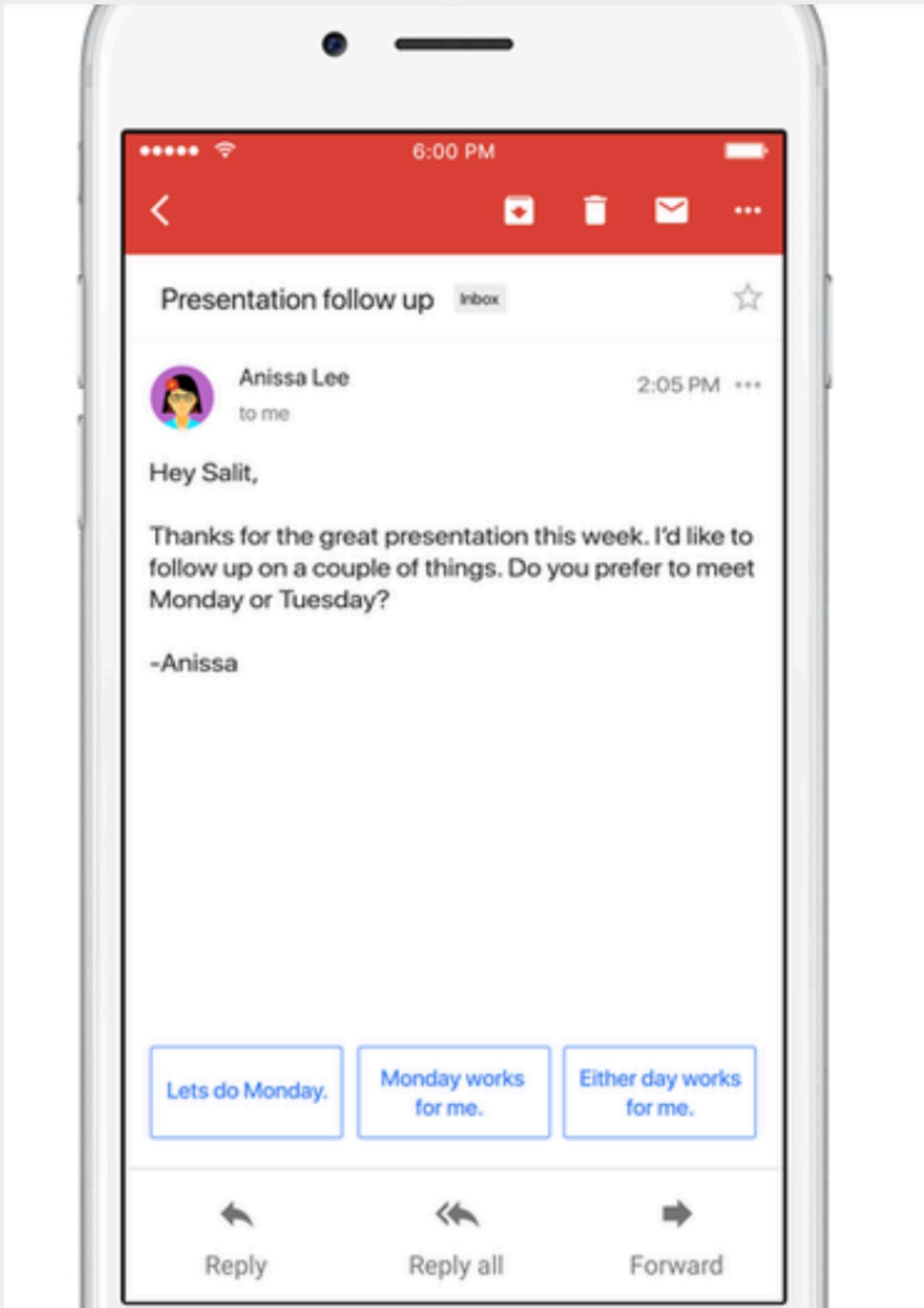
Enter Translation:
lebanese |
president
suffered
exposed
president emile
before
presented

5

T

Motivação:

- Algumas aplicações



T

Motivação:

Dan Jurafsky

Stanford University
Natural Language Processing

Language Technology

making good progress

mostly solved

Spam detection

Let's go to Agra! ✓
Buy V1AGRA ... ✗

Part-of-speech (POS) tagging

ADJ ADJ NOUN VERB ADV
Colorless green ideas sleep furiously.

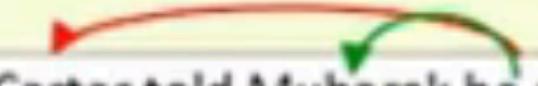
Named entity recognition (NER)

PERSON ORG LOC
Einstein met with UN officials in Princeton

Sentiment analysis

Best roast chicken in San Francisco! 
The waiter ignored us for 20 minutes. 

Coreference resolution

Carter told Mubarak he shouldn't run again. 

Word sense disambiguation (WSD)

I need new batteries for my **mouse**. 

Parsing

I can see Alcatraz from the window! 

Machine translation (MT)

第13届上海国际电影节开幕... 
The 13th Shanghai International Film Festival...

Information extraction (IE)

You're invited to our dinner party, Friday May 27 at 8:30  Party May 27 add

still really hard

Question answering (QA)

Q. How effective is ibuprofen in reducing fever in patients with acute febrile illness?

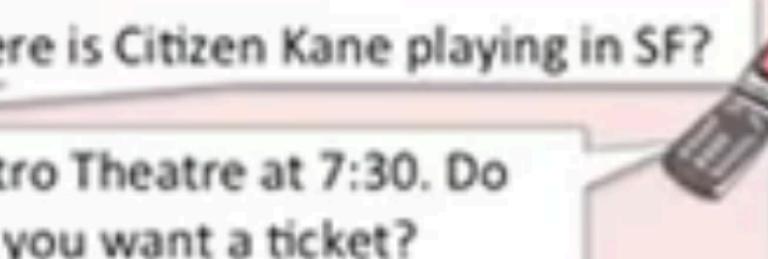
Paraphrase

XYZ acquired ABC yesterday
ABC has been taken over by XYZ

Summarization

The Dow Jones is up
The S&P500 jumped
Housing prices rose  Economy is good

Dialog

Where is Citizen Kane playing in SF?
Castro Theatre at 7:30. Do you want a ticket? 

Click with the mouse or tablet to draw with pen 4

O desafio:

- IMDB Movie Reviews

Criar um modelo que classifique um review em positivo ou negativo corretamente

