Introduction to NLP

What is Natural Language Processing?

Prof. Daniel Jurafsky, Stanford University

Presenter: Dr. Hamed Vahdat-Nejad

Dan Jurafsky

Question Answering

- IBM's Watson: **IBM Watson** is a computer system capable of answering questions posed in natural language. 2010-Now
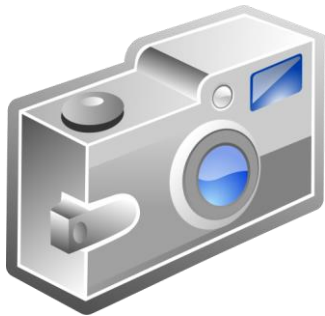
- ChatBot

- ChatGPT

Information Extraction

- CV processing to extract education, papers, etc.

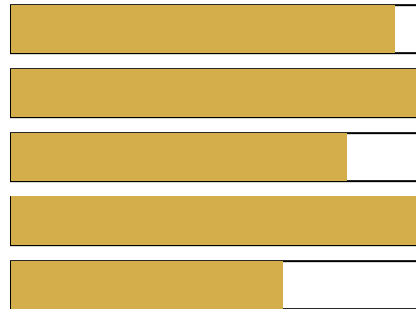**- Relevant Text extraction from big textual data**

Information Extraction & Sentiment Analysis

Attributes:

zoom

affordability

size and weight

flash

ease of use

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 fragile, is plastic and very light in weight you have to be very gentle in the handling of this camera

4

Machine Translation

- Fully automatic

- Helping human translators

Enter Source Text:

这 不过 是 一 个 时间 的 问题 .

Translation from Stanford's *Phrasal*:

This is only a matter of time.

Enter Source Text:

تعرض الرئيس اللبناني اميل لحود ل# حملة عنيفة في مجلس النواب الذي انعقد امس في جلسة تشريعية عادية تحولت الى " محاكمة " ل# رئيس الجمهورية علي موقف +ه من المحكمة الدولية و " الملاحظات " التي ادلي ب# +ها حول هذا الموضوع .

Translate   Clear

Enter Translation:

lebanese

president
suffered
exposed
president emile
before
presented
offer

Done!

5

Dan Jurafsky

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

Named-entity recognition is a subtask of information extraction that seeks to locate and classify named entities mentioned in unstructured text into pre-defined categories such as person names, organizations, locations, medical codes, time expressions, quantities, monetary values, etc.

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?

Ambiguity makes NLP hard

صنایع پیچیده ادبی و ابهام

100% REAL

Violinist Linked to JAL Crash Blossoms

Teacher Strikes Idle Kids

Red Tape Holds Up New Bridges

Hospitals Are Sued by 7 Foot Doctors

Juvenile Court to Try Shooting Defendant

Local High School Dropouts Cut in Half

Ambiguity is pervasive

New York Times headline (17 May 2000)

Fed raises interest rates   Fed raises interest rates 0.5%

Dan Jurafsky

Why else is natural language understanding difficult?

non-standard English

Great job @justinbieber! Were SOO PROUD of what youve accomplished! U taught us 2 #neversaynever & you yourself should never give up either♥

segmentation issues

the New York-New Haven Railroad
the New York-New Haven Railroad

idioms

dark horse
get cold feet
lose face
throw in the towel

neologisms

(واژه‌های جدید)
unfriend
Retweet
bromance

But that's what makes it fun!

Skills you'll need

- Simple linear algebra (vectors, matrices)

- Basic probability theory

- Java or Python programming
  - Weekly programming assignments