UNIVERSITI MALAYA
KUALA LUMPUR

INTRODCUTION TO NATURAL LANGUAGE PROCESSING

THEORY AND APPLICATIONS OF NATURAL LANGUAGE PROCESSING

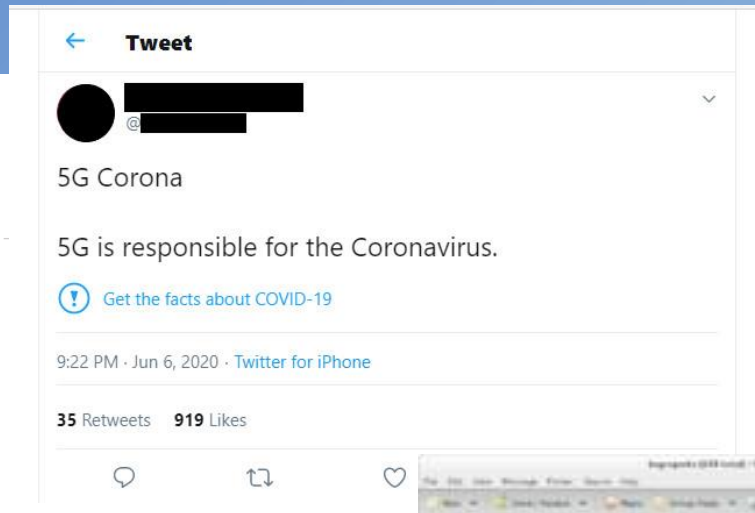
WOA7013

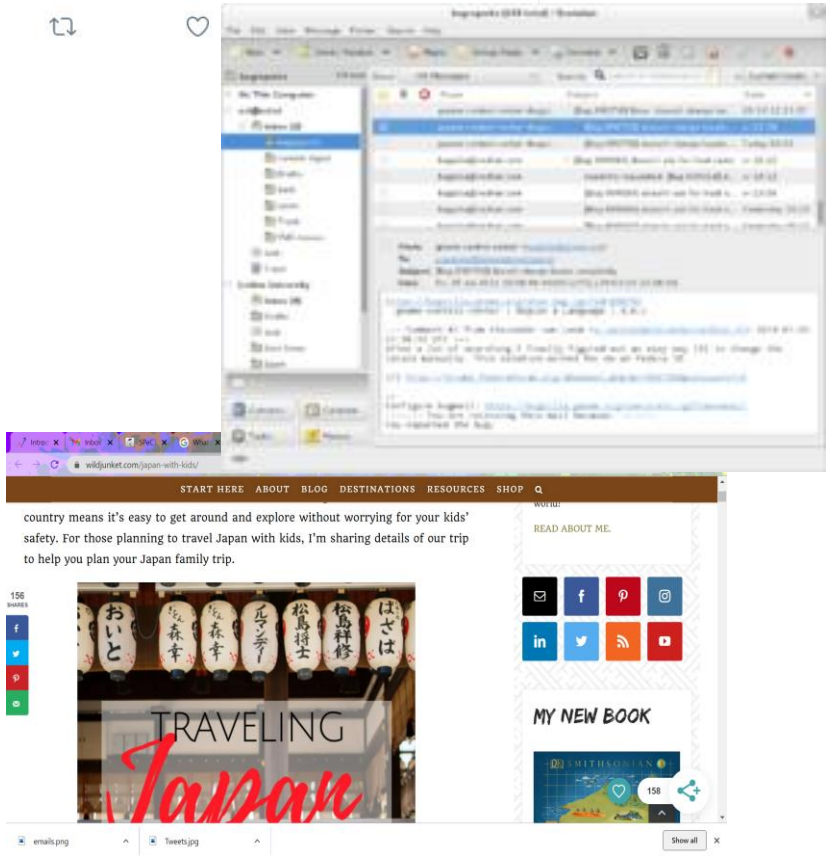
Language

- The ability to use language, perhaps more than any other attribute, distinguishes humans from other animals → one of human intelligence
- To represent the meaning of word, words have rich **meaning.**
- **Meaning:** the idea that is represented by a word or phrase.
- Human has gained a tremendous amount of experience in understanding natural language.

Data

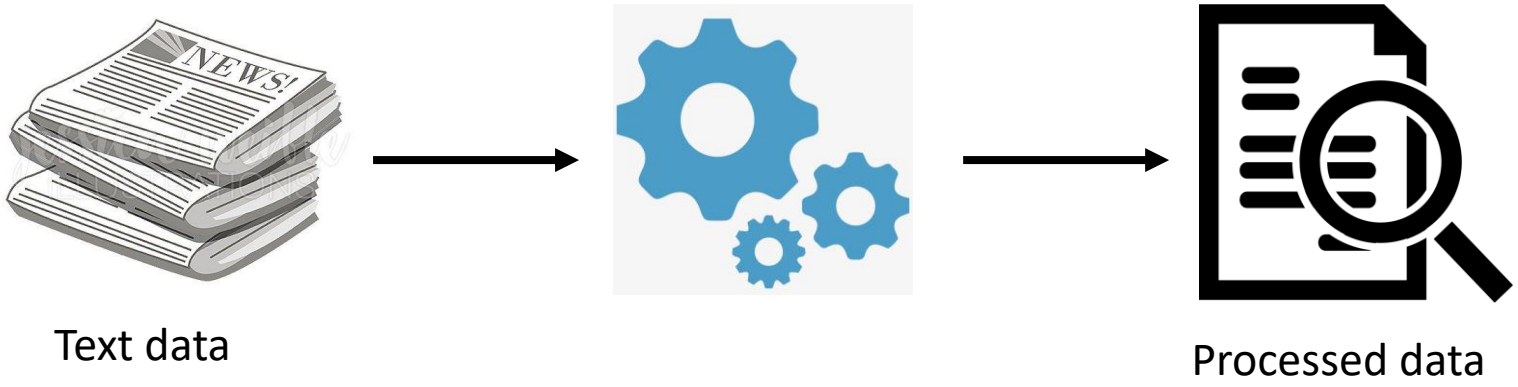
- Structured data
- Unstructured data:
    - i. Images, Video, audios and Text
    - ii. Text: the most unstructured form of all the available data
    - iii. Examples of Text:
        1. Social media data: tweets, posts, comments
        2. Conversation data: e-mails, chats
        3. Articles: news, blogs





Natural Language Processing (NLP)

- Extracting meaning from Text is hard.
- NLP is a sub-field of AI that is focused on enabling computers to understand and process human languages.
- Deriving useful information from text



History

- WARREN WEAVER had first mentioned the possibility of using digital computers to translate documents between natural human languages in March 1947.
- **Machine translation** is one of the earliest applications of computer
- A computer translation based on expertise in breaking enemy codes during World War II
  - Major attempts in US and USSR
  - Russian to English and reverse

Chatbot: ELIZA

Early NLP Systems

- https://www.chatbots.org/chatbot/eliza
- was developed in the mid-1960s at MIT by Joseph Weizenbaum.
- replicates the conversation between a psychologist and a patient, simply by permuting or echoing the user input.
- Eliza simulated conversation by using a *pattern matching* and *substitution methodology* to demonstrate the communication between man and machine and creating a simple illusion of understanding.

- How it works?
  - It has keywords database where for each keyword, the system stores an integer, a pattern to match against the input and a specification of the output.
  - Algorithm: Given a sentence *S*, find a keyword in *S* whose pattern matches *S*. If there is more than one keyword, pick the one with the highest integer value. If there are no keywords, generate a continuation statement, such *as Tell me more* and *Go on*.
- Sample data from ELIZA:

Word	Rank	Pattern	Outputs
alike	10	?X	In what way? What is the connection, do you suppose?
are	3 3	?X are you ?Y ?X are ?Y	Would you prefer it if I weren't ?Y? What if they were not ?Y?

Why NLP is hard?

It's the nature of the human language that makes NLP difficult.

- Highly ambiguous at all levels
- Complex and subtle use of context to convey meaning
- True understanding what a piece of text means in the real-world.

For example, what do you think the following piece of text means?

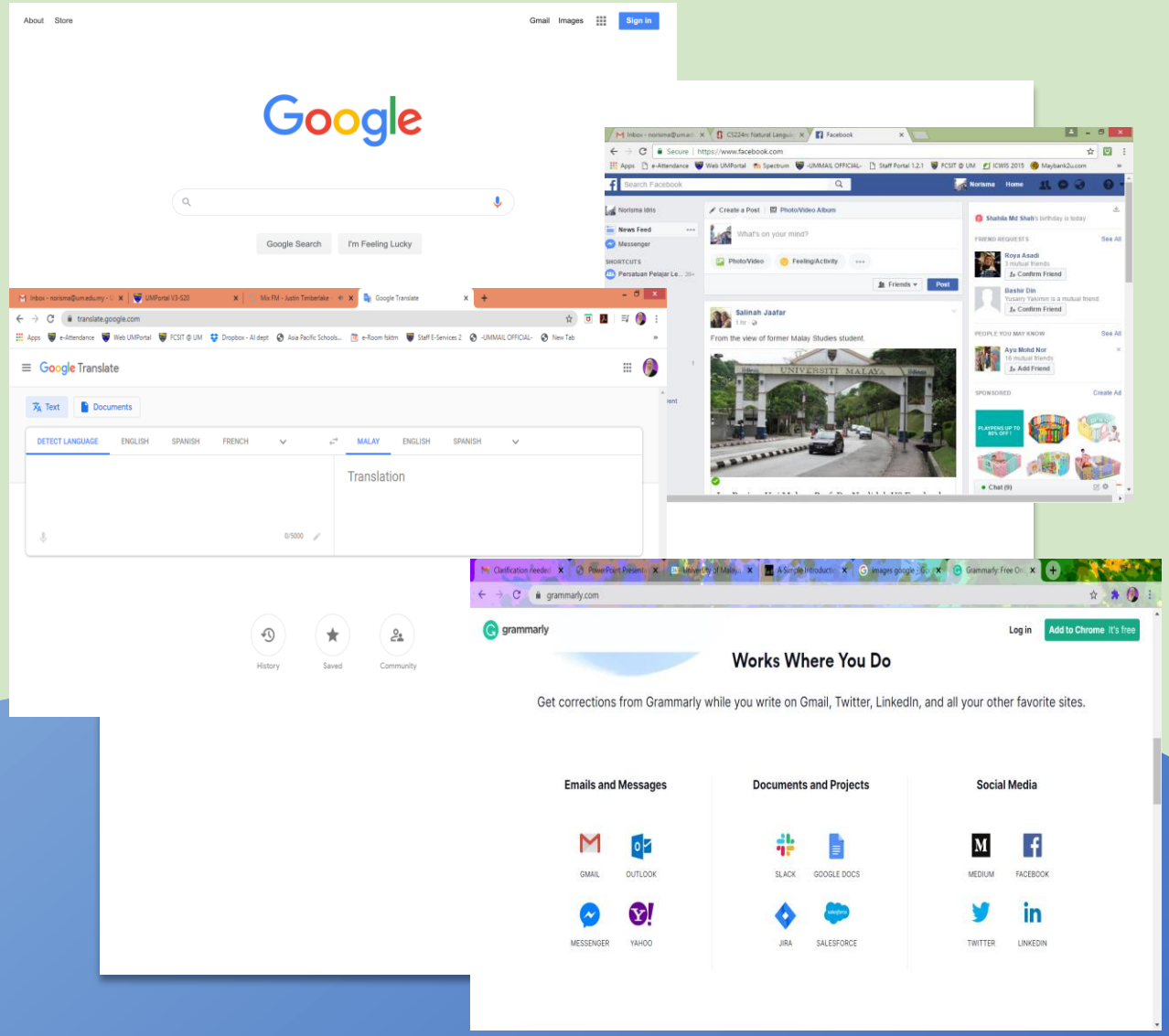
- *"Lionel Messi was on fire last night. He totally destroyed the other team."*
- *"Minister Accused of Having 8 Wives in Jail" (headlines)*


“It’s okay if you don’t like me.
Not everyone has good taste.”
Parade

What is NLP used for?

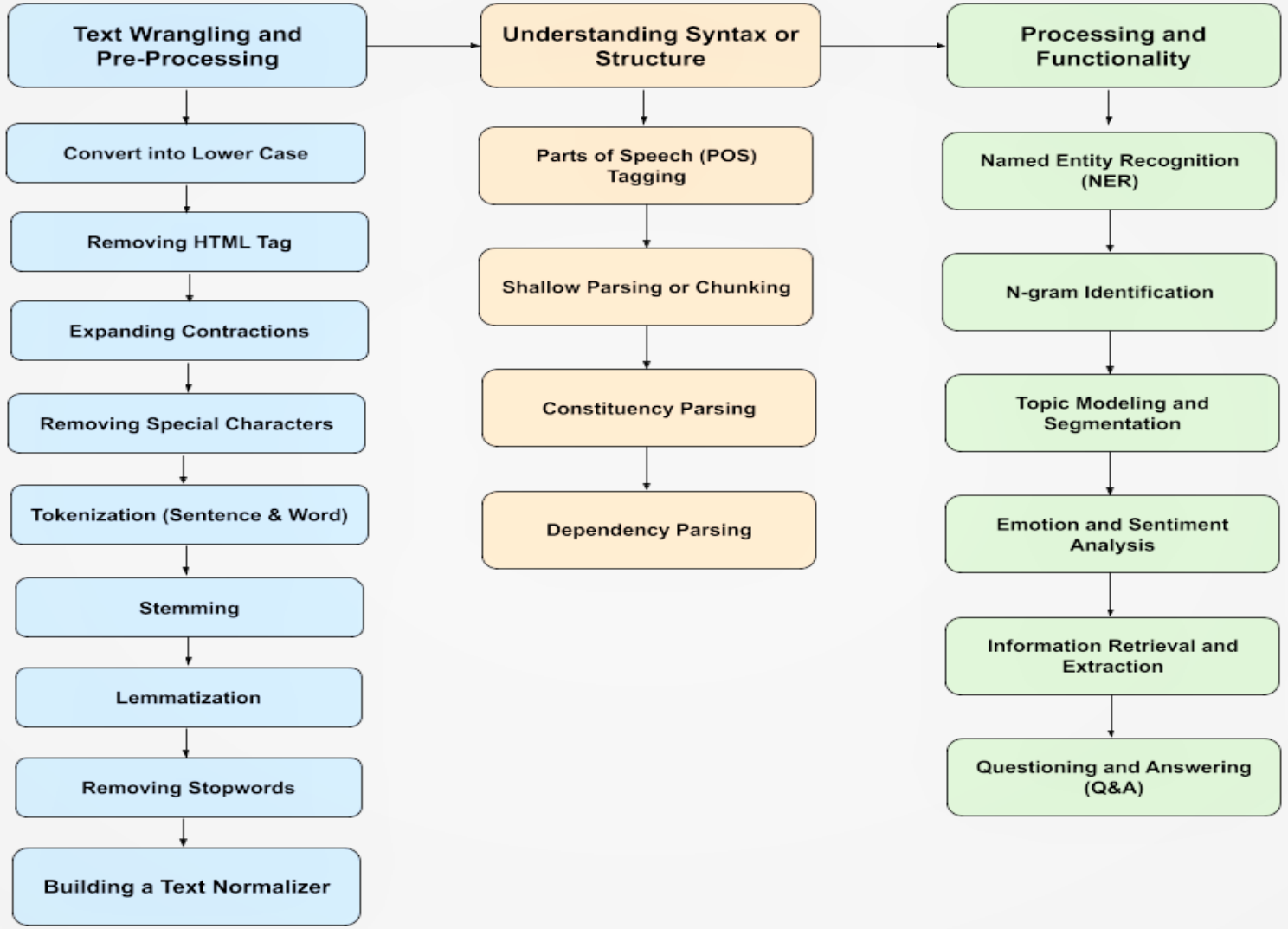
Applications of NLP are everywhere because people communicate most everything in language:

- Web search
- Translation applications
- Word Processors
- Personal assistant
- Social Medias



More and more companies are applying natural language processing applications to everyday business problems.

NLP Pipeline



https://medium.com/@suneelpatel.in/nlp-pipeline-building-an-nlp-pipeline-step-by-step-7f0576e11d08

What can NLP do?

Google Duplex: A.I. Assistant Calls Local Businesses To Make Appointments

Conclusion

NLP is already used in many systems today.

Many technologies are in use, and still improving.

Lots of awesome research to work on!

Language Technology

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

MAKING GOOD PROGRESS

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

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?

Sources: Stanford University