NATURAL LANGUAGE PROCESSING (NLP) FUNDAMENTALS

INTRODUCTION

DR LEK HSIANG HUI

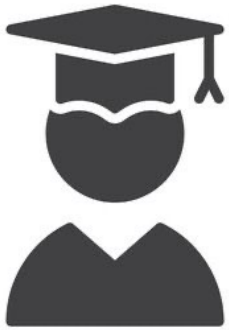
About Me

- LEK Hsiang Hui <hsianghui@nus.edu.sg>
- Senior Lecturer
- Department of Information Systems and Analytics

- Teaches Courses on:
  - Analytics
  - Software Engineering

- Undergraduates, Graduates, and Executives

About Me



Did PhD in the area Natural Language Processing (Sentiment Analysis)

Contact

https://www.facebook.com/hsianghui

@hsianghui

https://www.linkedin.com/in/hsianghui/

hsianghui@nus.edu.sg

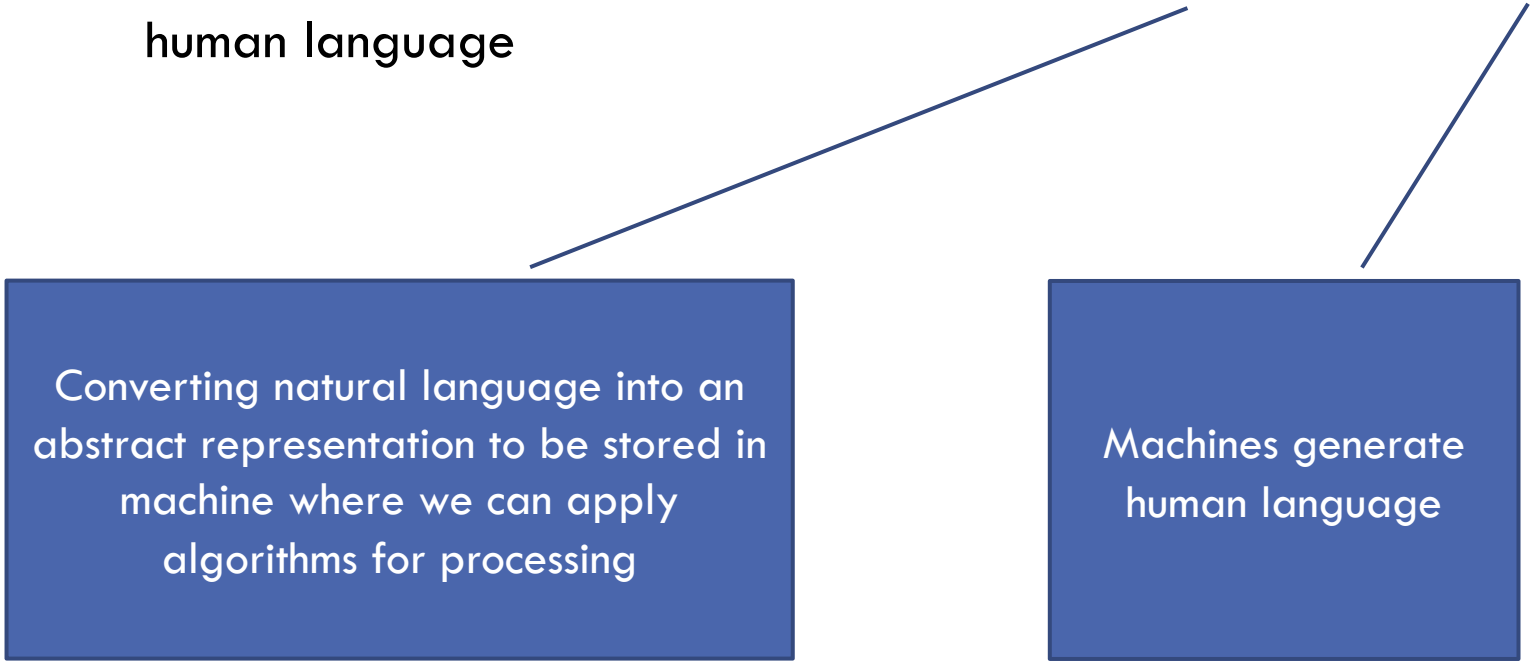
Introduction to NLP

What is NLP?

- Natural Language = Human Language
  - Can be in the form of text and speech
- Natural Language Processing (NLP)
  - Subfield of linguistics, computer science, and artificial intelligence
  - Focusing on **giving computers the ability to understand human language**
  - Will focus on text for this course

What is NLP?

- Another definition:
  - Application of computational techniques to **analyze** and **synthesize** human language



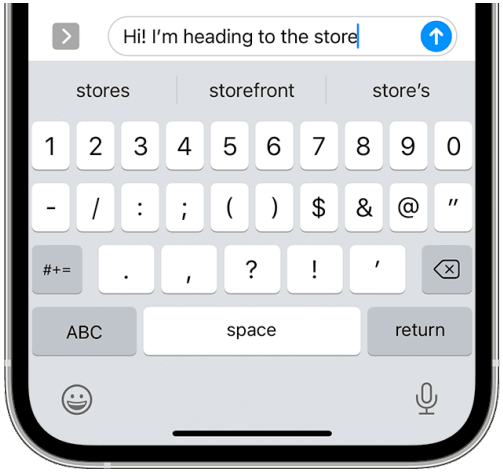
Why NLP?

- Theoretically speaking:
  - Natural language understanding has been a major goal of AI Research
- Practically speaking:
  - Many practical applications that we see in our lives

Siri

Google Translate

Spelling & Grammar Correction



Next/Current Word Prediction

Applications of NLP

Applications of NLP

- Information Extraction (IE)

Turnpike Designs Co.

INVOICE

Turnpike Designs
156 University Ave, Toronto
ON, Canadam M5h 2h7

416-555-1212

BILL TO
Jiro Doi
1954 Floor Street West
Toronto, ON, M6P 3K9
Canada

J_doi@example.com
416-555-1212

Invoice Number: 14
PO./S.O. Number: AD29094
Invoice Date: 2018-09-25
Payment Due: Upon receipt

Amount Due (USD): $2,608.20

Services	Quantity	Price	Amount
Platinum web hosting package Down 35mb, Up 100mb	1	$65.00	$65.00
2 page website design Includes basic wireframes, and responsive templates	3	$2,100.00	$2,100.00
Mobile designs Includes responsive navigation	1	$250.00	$250.00

Subtotal: $2, 145.00
Tax 8%: $193.20

Total: $2, 608.20

Amount due (CAD): $2, 608.20

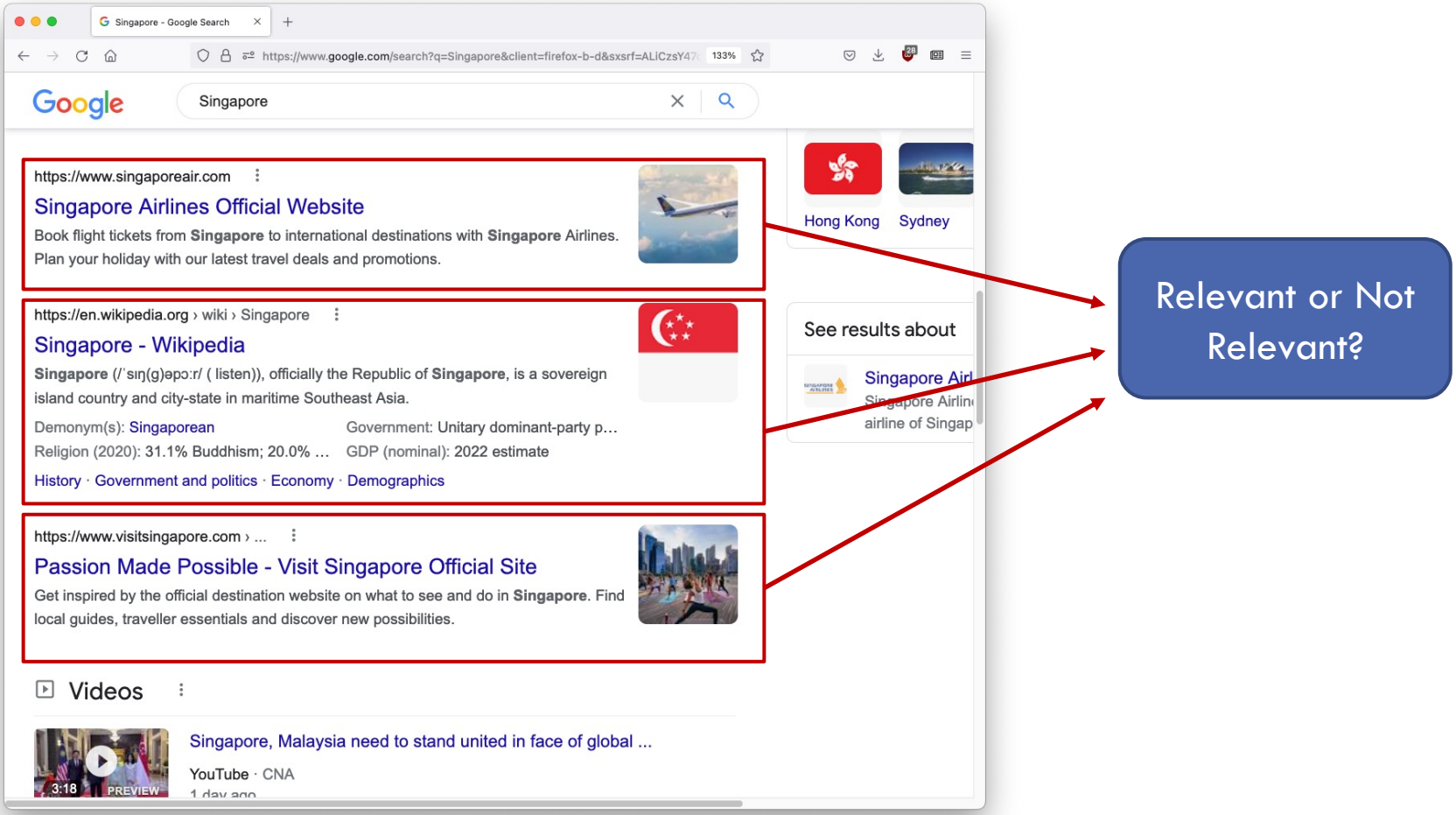
Source: https://nanonets.com/blog/information-extraction/

Applications of NLP

- How do we achieve Information Extraction (IE)?
  - **Rule-based**
    - Learn extraction rules
  - **Classification-based**
    - Use machine learning techniques
    - Treat the IE task as a classification task
    - Assign classes to each token
    - Combine consecutive tokens into entities

Applications of NLP

- Information Retrieval (IR)

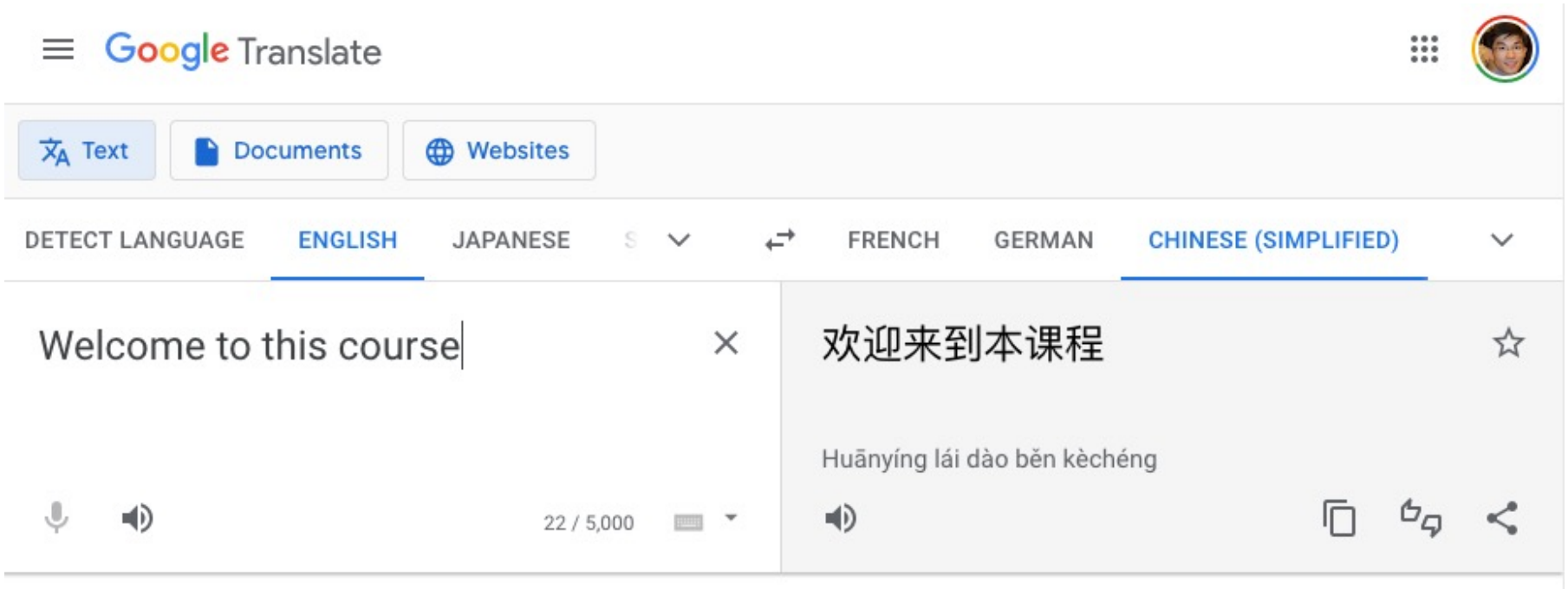


Applications of NLP

- How do we achieve Information Retrieval (IR)?
  - **Vector Space Model** Approach
    - Represent documents and query as vectors
    - Calculate the cosine similarity between the vectors
    - Vector of document should be similar to the vector of query
    - Rank the documents based on the similarity value between document and query and choose top n most similar documents

Applications of NLP

- Language Detection & Machine Translation (MT)



Applications of NLP

- How do we achieve Machine Translation (MT)?
  - **Rule-based**
    - Make use of many linguistic rules and bilingual dictionaries for each language pair
  - **Statistical-based**
    - Make use of multilingual corpora
    - E.g. European parliament minutes are written in all European Union (EU) official languages
    - Does not need to know the grammar

Applications of NLP

- Text Summarization

World

EU to propose price cap on Russian gas, von der Leyen says

07 Sep 2022 11:37PM
(Updated: 07 Sep 2022 11:40PM)

European Commission President Ursula von der Leyen speaks during a news conference on the energy crisis, in Brussels, Belgium, on Sep 7, 2022. (Photo: REUTERS/Johanna Geron)

BRUSSELS: The European Commission will propose a price cap on Russian gas, alongside measures including a mandatory EU cut in electricity use and a ceiling on the revenue of non-gas power generators, the bloc's chief said on Wednesday (Sep 7).

Power and fuel prices have soared as Russia has cut the amount of gas it sends to Europe, with European leaders accusing Moscow of weaponising its energy supplies. Russia has blamed the supply cuts on technical issues and Western sanctions imposed over its invasion of Ukraine.

"We will propose a price cap on Russian gas ... We must cut Russia's revenues which Putin uses to finance this atrocious war in Ukraine," European Commission President Ursula von der Leyen told reporters.

President Vladimir Putin said on Wednesday that Russia will stop supplying gas and oil if price caps are imposed.

Some EU countries - which would need to approve the EU proposals - are wary of a cap on Russian gas prices if it puts at risk the dwindling supply they still receive from Moscow.

Also worth reading

Malaysia lifts indoor mask mandate, excludin...

Commentary: Shutting down Nord Stream 1...

Phoon Chiu Yoke seeks court's green light to...

Najib issued cheques totalling RM6.4 million t...

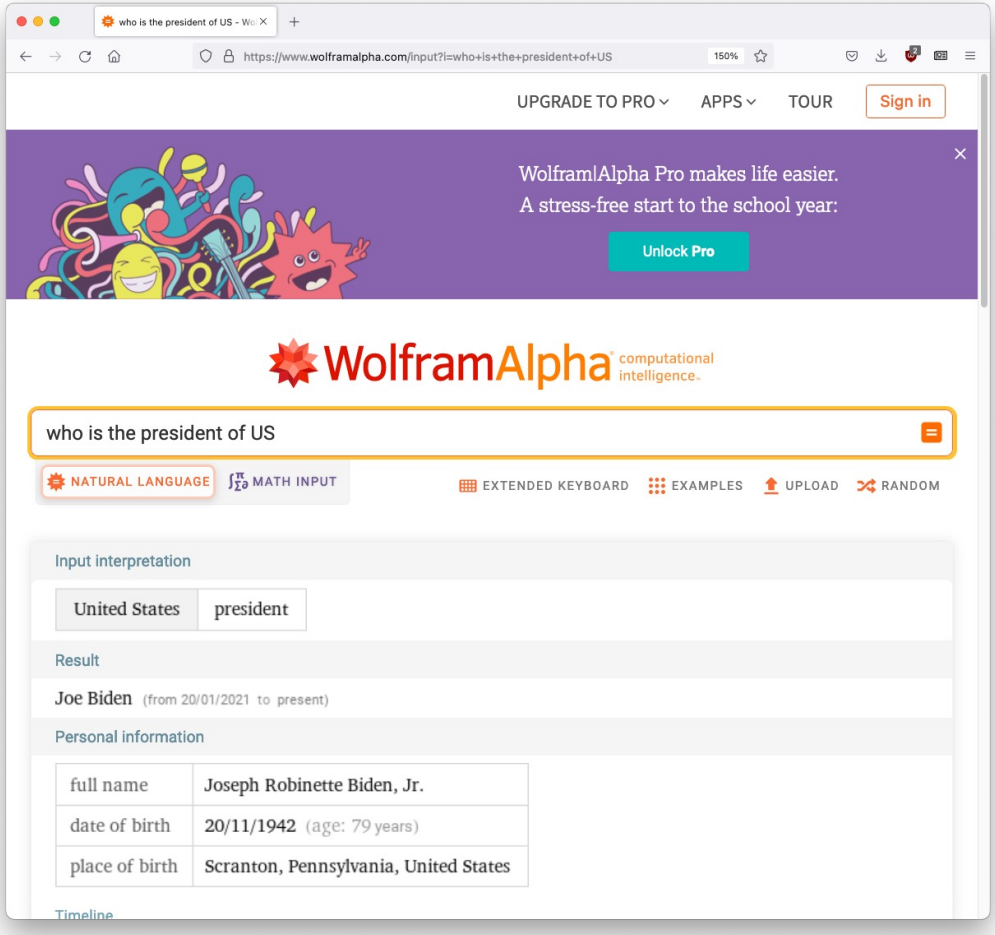
The European Commission will propose a price cap on Russian gas, alongside measures including a mandatory EU cut in electricity use and a ceiling on the revenue of non-gas power generators. Power and fuel prices have soared as Russia has cut the amount of gas it sends to Europe. EU energy ministers will discuss emergency plans to lower gas and power prices on Friday. Germany's Energy Minister Ursula von der Leyen said the EU wants to cap the revenue of non-gas fuelled generators and channel "unexpected profits" into measures to support households and businesses.

Applications of NLP

- How do we achieve Text Summarization?
  - **Extractive Text Summarization**
    - Identify the significant sentences of text and combine them into a summary
  - **Abstractive Text Summarization**
    - Interpret the context and generate new sentences from the original content

Applications of NLP

- Question Answering (QA)

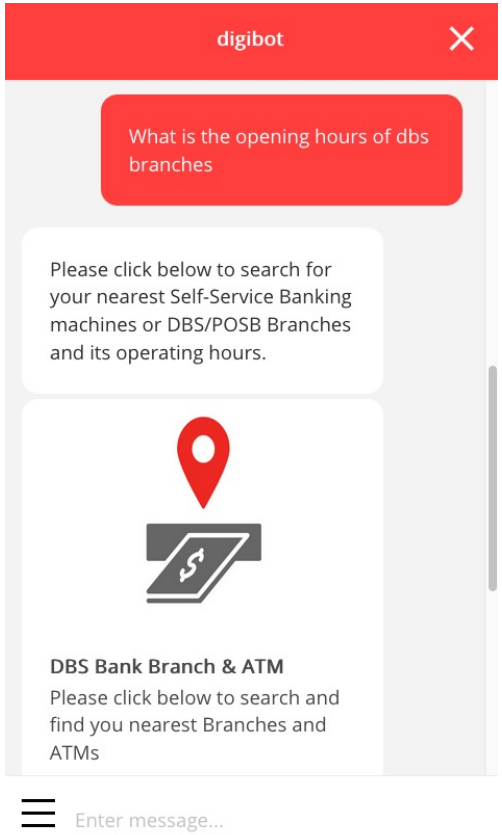


Applications of NLP

- How do we achieve QA?
  - Consist of many components such as:
    - Question Parser
    - Query Generator
    - Searcher
    - Answer Candidates Extractor
    - Answers Ranker

Applications of NLP

- Conversational Agents (e.g. Chatbots)



Applications of NLP

- How do we build a chatbot?
- Depends what the chatbot is used for:
  - If meant to be used for answering questions, essentially it is similar to QA
  - If meant to be used to maintain a conversation with the user, need to be able to connect **multiple chat messages** and build up **context**
- The major NLP task is **intent classification**
  - i.e. understanding the intent of an input message

Applications of NLP

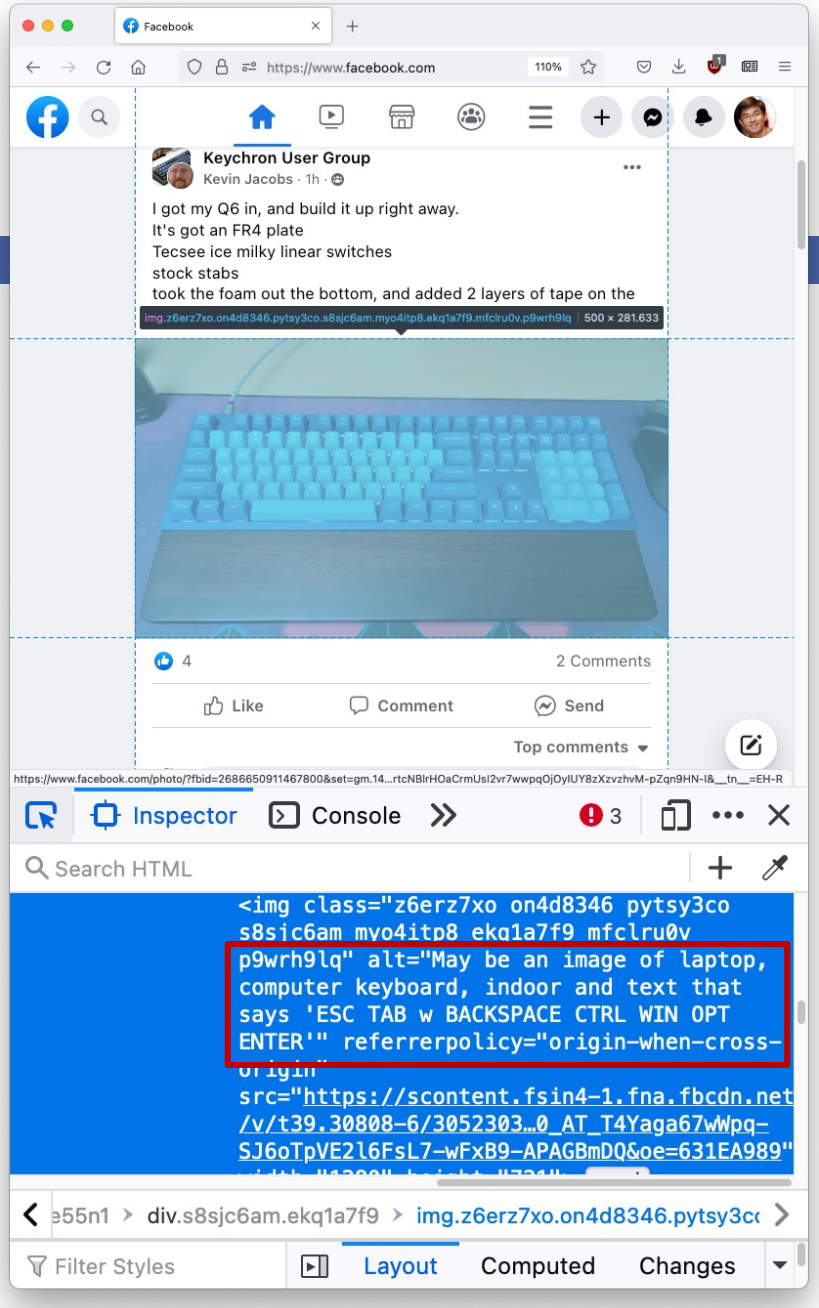
- Document/Text Classification
  - Spam Detection
  - Categorizing the topic of a document
  - Sentiment Analysis

Applications of NLP

- How do we achieve document/text classification?
  - Need to convert text (which is in unstructured form) into a structured form
    - **Feature Engineering**
    - **Tokenization**
    - **Feature Vector Generation**
  - Apply Machine Learning classification on the **feature vector**
- Will cover this in more detail in the next course (Sentiment Analysis Fundamentals)

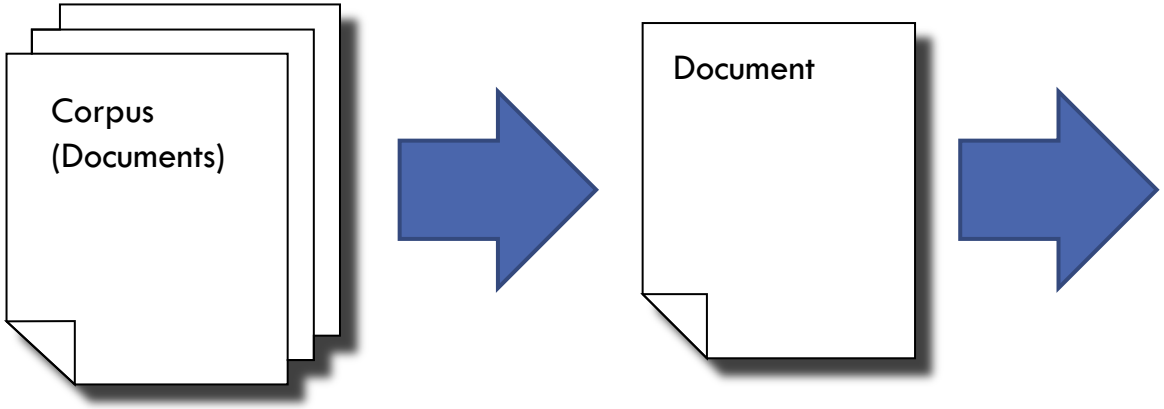
Applications of NLP

- Image Captioning
  - Uses both NLP and **Computer Vision** to generate the captions



Language Processing

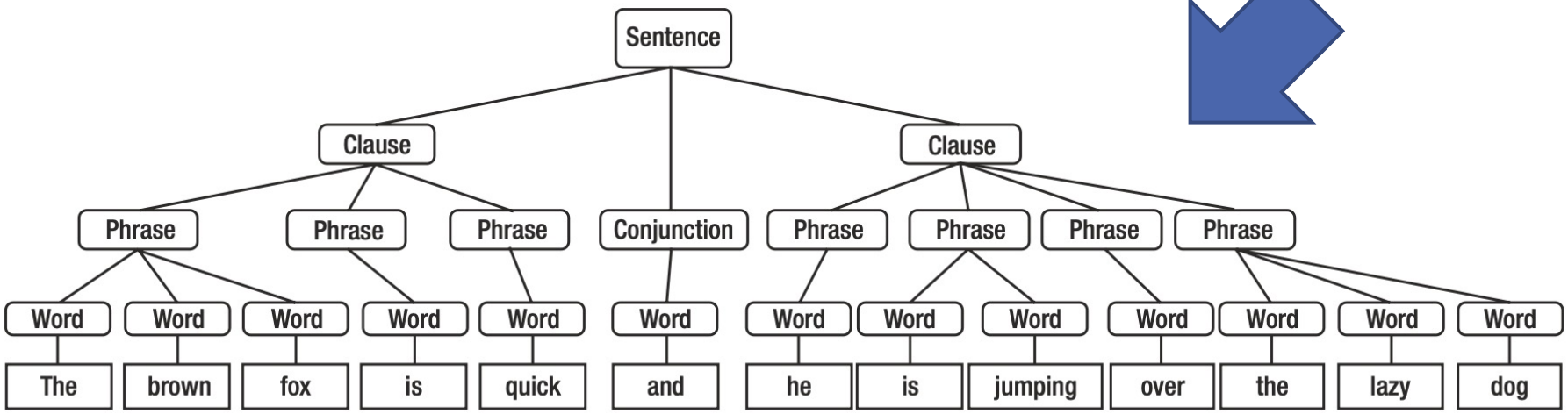
Language Processing



Ut inventore dolores aut error commodi et optio magni et dolorum quidem a error velit et numquam quaerat est unde ullam! Quo excepturi quae consequuntur ipsa aut minima quod nam sunt quasi ut ipsa velit aut voluptate neque ex magnam atque. Et quae obcaecati et voluptatem voluptas ad iusto eaque et harum quia. Id saepe voluptate sit sunt autem ea porro odio est necessitatibus nisi.

Est atque rerum et dolor nemo qui magni incidunt. Aut natus perspiciatis sit corrupti ullam eum ipsum eius est debitis voluptatem.

Ratione molestiae aut itaque corporis et minus dolore! Aut dicta quam id corporis nihil non voluptatem galisum. Non eius sunt sit Quis nobis est deserunt commodi. Et eaque quae qui officiis dolor aut corrupti ipsa aut consequuntur alias sed esse deserunt doloribus minima ut doloremque laboriosam.



Types of Analysis

Types of Analysis

1. **Lexical Analysis**
   - Understanding the structure & meaning of words
2. **Syntactic Analysis**
   - Organization of words based on their structure/ordering in the sentence
3. **Semantic Analysis**
   - Meaning of words, phrases and sentence
4. **Discourse Analysis**
   - Works with multiple consecutive sentences to interpret things such as pronouns
5. **Pragmatic Analysis**
   - Understanding and interpretation of language in real-world context

1. Lexical Analysis

Lexical Analysis

- *"Understanding the structure & meaning of words"*
- Core Concepts:
  - **Sentence Segmentation**
  - **Tokenization**
  - **Text Normalization** (e.g. Stemming , Lemmatization)

2. Syntactic Analysis

Syntactic Analysis

- "Organization of words based on their structure/ordering in the sentence"
- Core Concepts:
  - **Part-of-Speech (POS) Tagging**
  - **Shallow Parsing (Chunking)**
  - **Dependency Parsing**

3. Semantic Analysis

Semantic Analysis

- *"Meaning of words, phrases and sentence"*
- Core Concepts:
  - **Word Sense Disambiguation (WSD)**
  - **Named Entity Recognition (NER)**
  - **Semantic Role Labeling (SRL)**

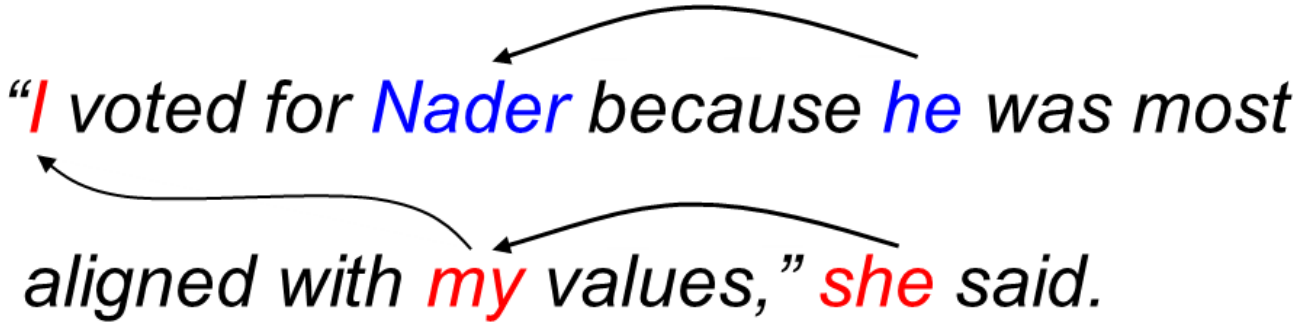
4. Discourse Analysis

Discourse Analysis

- *"Works with multiple consecutive sentences to interpret things such as pronouns"*
- Core Concepts:
  - **Coreference Resolution**
  - **Ellipsis Resolution**

Coreference Resolution

- Coreference Resolution
  - Task of identify expressions that refer to the same entity in text
  - Entities can be named entities, noun phrases, pronouns, etc



Ellipsis Resolution

- Ellipsis Resolution
  - Task of inferring the ellipses using the surrounding context
  - Ellipsis = … (words appear to be missing)

Lauren can play something, but I don't know what

Lauren can play something, but I don't know what Lauren can play

Lauren can play the guitar and Mike can, too

Lauren can play the guitar and Mike can play the guitar, too

5. Pragmatic Analysis

Pragmatic Analysis

- *"Understanding and interpretation of language in real-world context"*
- Core Concepts:
  - **Textual Entailment**
  - **Intent Classification**

Textual Entailment

- Textual Entailment
  - Task of determining the directional relation between text fragments
  - Given text ***t*** and hypothesis ***h***, ***t*** entails ***h*** (***t*** ⇒ ***h***)
    - A human reading *t* would infer that *h* is most likely true

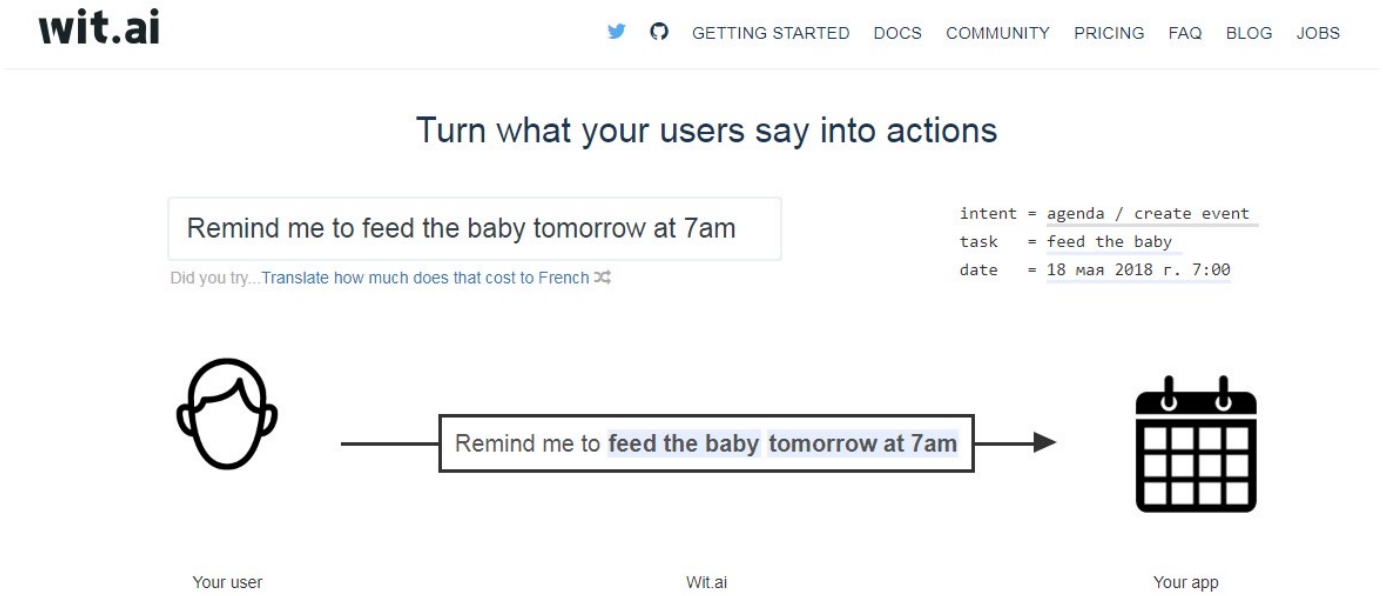
t: If you help the needy, God will reward you

⇒

h: Giving money to a poor man has good consequences

Intent Recognition

- Intent Classification
  - Task of determining the intent of the written/spoken input
  - Recall: this is a major task in chatbots



NATURAL LANGUAGE PROCESSING (NLP) FUNDAMENTALS

INTRODUCTION

DR LEK HSIANG HUI