

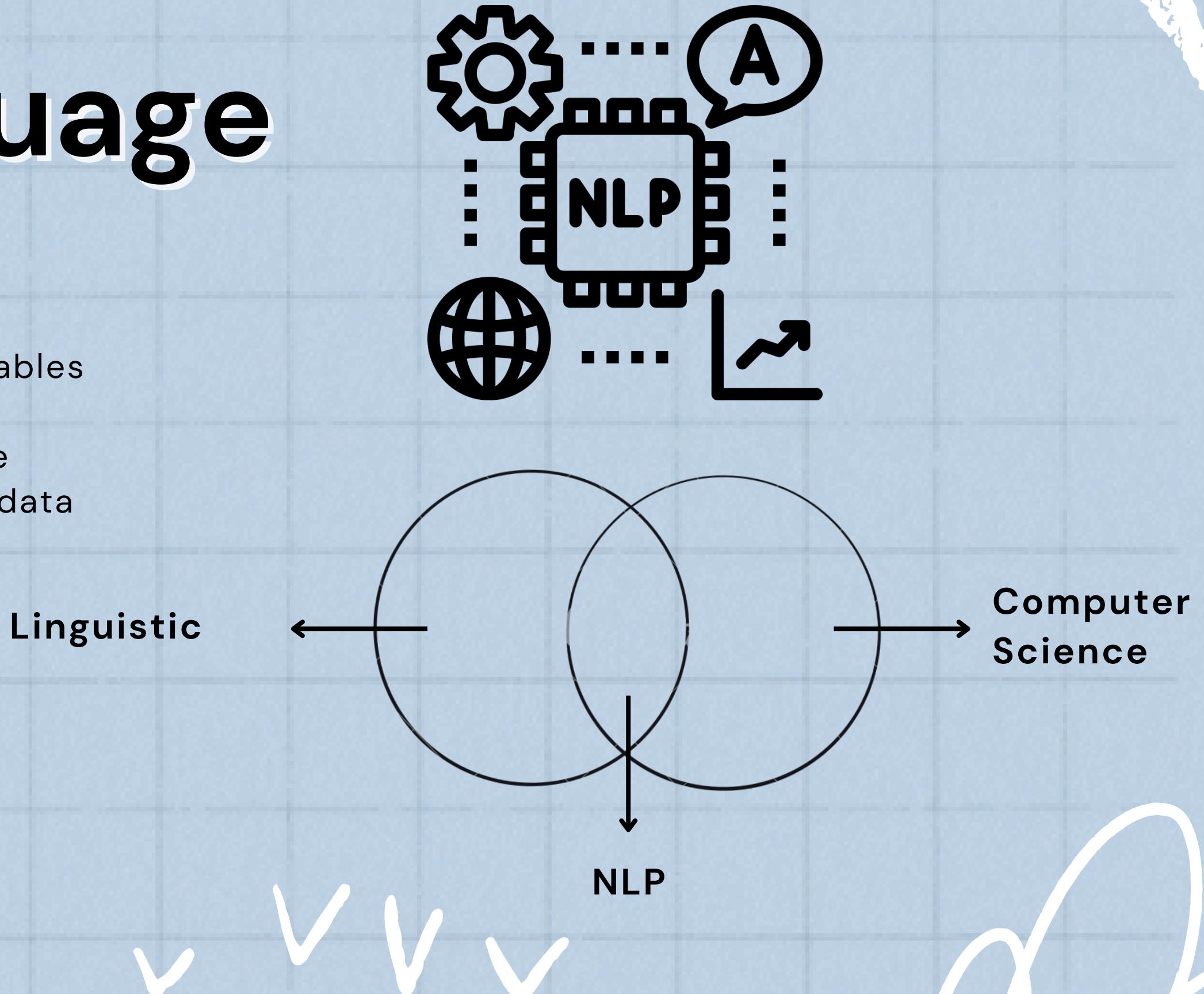
SS234522 - Data Mining and Visualization - Week 10 - Practical Session

Natural Language Processing

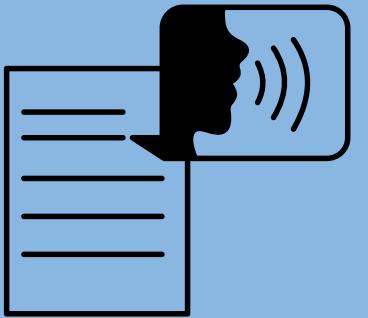
Assistant : Abraham Sumarsono – 5003201099

Natural Language Processing

A branch of artificial intelligence (AI) that enables computers to comprehend, generate, and manipulate human language. Natural language processing has the ability to interrogate the data with natural language text or voice.

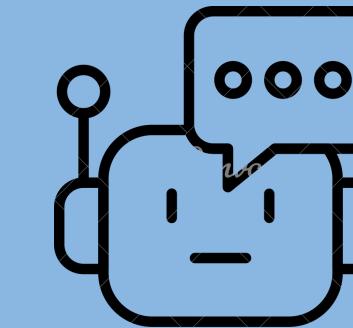


Speech-to-text



- Allows machines to understand and read human language with the intent to respond and react like people do.
- Steps: uploading audio data -> feature engineering -> converting signal into digital data -> transcribing audio data

Chatbots and Virtual Assistants



- To converse with human customers or clients and solve or understand their problems before a human executive can intervene
- Are pre-programmed with potential inquiries and responses
- Example: Siri, Google Assistants, Duolingo

NLP Use Cases

Exploring data mining functionalities

Sentiment Analysis



- Used to categorize the nuance of sentences—positive, negative, or neutral
- Mostly used to monitor public's overall sentiment in social media conversations regarding your brand or product

Predictive text



- Predicting what users are going to say in through their keyboard
- Example: smart predictions in your keyboard!

The Steps

Segmentation

Cricket was invented in England,
supposedly by shepherd who herded
their flock

Cricket was invented in England
supposedly by shepherd who herded
their flock

Tokenizing

Cricket was invented in England

Cricket was invented
 in England

Stemming

Eliminates prefixes and suffixes from words, transforming them into **their fundamental or root form**

Listen + s

Listen + ed

Listen + ing

Lemmatization

a text pre-processing technique used in natural language processing (NLP) models to break a word down to its **root meaning to identify similarities.**

Better

Good

Worse

Bad

Stop Words

Cricket was invented in England

Example of more stop words:
a, of, are, the, on, to

Speech-Tagging

Cricket was invented in England

Noun

Verb

Noun

Noun

Preposition

Natural Language Toolkit

A leading platform for building Python programs to work with human language data. It provides easy-to-use interfaces to over 50 corpora and lexical resources such as WordNet, along with a suite of text processing libraries for classification, tokenization, stemming, tagging, parsing, and semantic reasoning, wrappers for industrial-strength NLP libraries, and an active discussion forum.



**Thank you
very much!**