

Natural Language Processing:

1.

Written Language

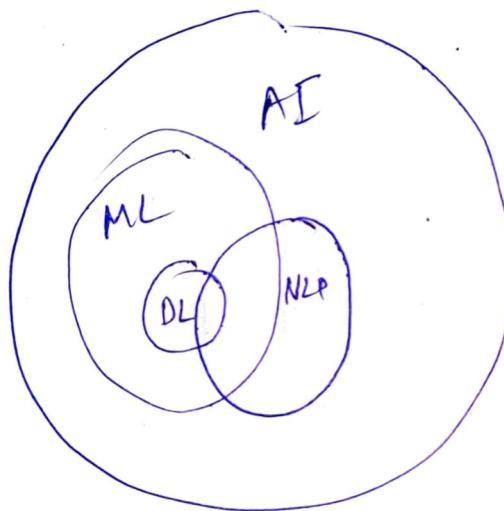
- ↳ easier in processing
(unambiguous word)

2.

Spoken Language

- ↳ Harder to interpret

AI and NLP (Difference and Commonness)



is a subfield of linguistics,
computer sci and AI,
concerned with the interaction
between computer and
Human language

Applications of NLP in Modern Era:

7 Major Applications of NLP -

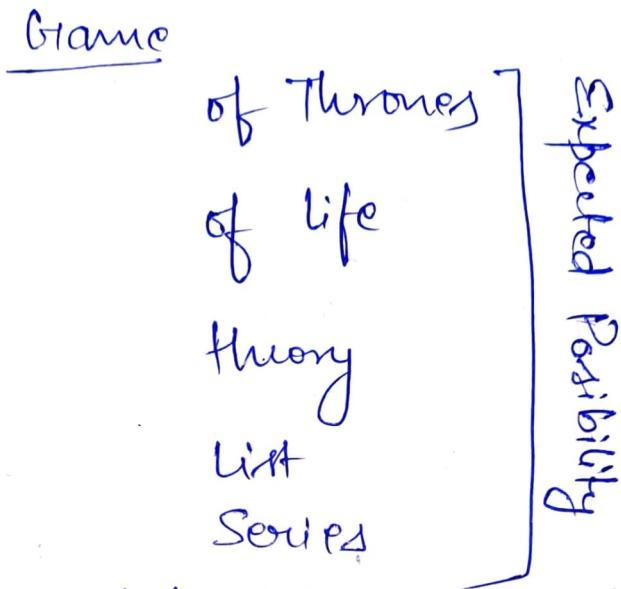
① Chatbots: They are form of AI that are programmed to interact with humans in such way that they sound like human themselves

↳ Simple - Based on specific keyword

↳ Complex and interactive - Have knowledge base that can respond with experience.

2. Auto Complete in Search Engines: It guesses the next keyword, you will be typing.

Example -



Uses enormous data sets, to analyze what their user/customer want to type.
It finds the most probably, frequent word that suggest more common possibilities.

It finds the connection that make sense and How they are interconnected to each other.

Data Set

Indexing

Crawler

Suggested words

Analyzing with data

③ Voices Assistants: Siri, Alexa, Google Assistant ②

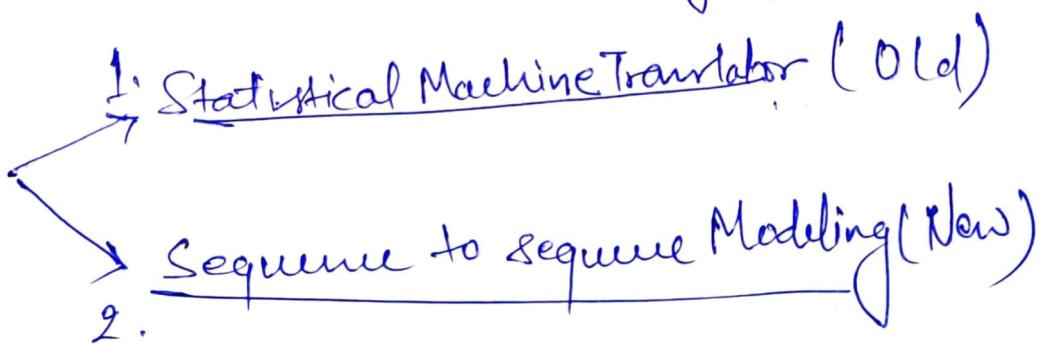
- Task → Makes Call
→ Schedule a calendar
→ Schedule Meeting
→ Set Alarms
→ Surfing Internet
→ Shopping
- Greeting and sending text msg
→ Managing smart Home appliances

They uses complex combination of speech recognition.
→ Natural Language Understanding

→ NLP

Drawback - They are not understanding all the time.

④ Language Translator: tool to convert text from one language to another language



In Sequence to Sequence , the algorithm to convert a sequence of words from one language to another language.

In Statistical Machine Translation \Rightarrow they analyze millions of documents that were already translated from one language to another , it looks for common pattern, and basic vocabulary of language.

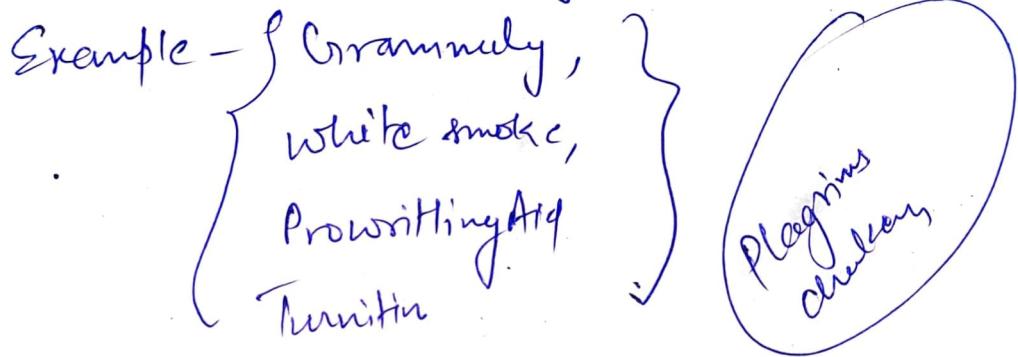
(5) Sentiment Analysis:- Company uses sentimental analysis in social media applications , to know how a particular type of user feel about a particular topic or user.

They - NLP, computational linguistics, text analysis etc.

- Uses - Product Advertisement
- Threats to national security
- Happy, sad, neutral emotions towards something
- Branding and getting public opinion
- 

6. Grammar Checkers — important tool for any professional writer. ③

- They can find errors
- Suggest better option
- Suggest better synonyms
- Improve readability of sentence



⑦ Email Classification and Filtering

→ Primary
→ Social
→ Promotions

Uses NLP to identify the content of each mail with text classification so that it can be put in correct section

Phishing and Filtering

Monitoring

Phases of Natural Language Processing:

Processing written text - lexical, syntactic and semantic knowledge, and real world information.

Processing Spoken language: using all the function of processing written text plus additional information about phonology as well as information to handle the further ambiguities that rises in the speech.

Problem and Solution Sentence with book → - 206

Phases $\not\rightarrow$ NLP \Rightarrow

- ① Morphological Analysis
- ② Syntactic Analysis
- ③ Semantic Analysis
- ④ Discourse Integration
- ⑤ Pragmatic Analysis

- ① Morphological Analysis: Individual words are analyzed into their component.
- ② Syntactic Analysis: linear sequence of word transformed into structure

~~Boy the go the to store~~

Boy the go the to store

will be rejected by the analyzer

- ③ Semantic Analysis: It assigns meaning mapping is made between the syntactic structure and object in the task domain.

If no mapping is possible then it may be rejected.

- ④ Discourse Integration - Meaning of individual sentences may depend on the sentence that precedes it or may be influenced by the meaning of it.

- ⑤ Pragmatic Analysis - The structure representing what was said is reinterpreted to determine what actually meant Do you know what time it is it has to request to tell time.

Example: I want to print Bill's .init file. (5)

① Morphological Analysis:

① Bill's → into proper noun,
's → possessive suffix "s"

② Recognize the sequence ".init" as a file extension
that is functioning as an adjective in the
sentence.

→ This process will provide or usually assigns
syntactic categories to all the words in sentence

② Syntactic Analysis: Must exploit the
result of morphological analysis to build structural
description of the sentence.

The goal is process, called parsing.

Parsing: is to convert the flat list of words
that forms the sentence into structures that define the
units that are represented by that flat list

③ Semantic Analysis -

Two important part

- mapping the individual word into appropriate object in the knowledge base
- create correct structure

How to build an NLP pipeline:

① Sentence segmentation: first step for building NLP Pipeline, it breaks the paragraph into separate sentences.

Step-2: Word Tokenization: Used to break sentence into ~~so~~ separate words

Step-3: Stemming: Used to normalize words into base form or root form. (Root word does not have Meaning)
 eat ate eaten → will belong to one same tree.

Step 4: Lemmatization: similar to stemming used to group inflected form of the words, called lemma.

The root word has meaning in lemmatized

Step 5: Identifying Stop words →

there are a lot of words that appears very frequently is, and, the and a

these words are flagged in statistical analysis.

Step-6: Dependency Parser: used to find how all the words in the sentence are related to each other.

Step-7: POS Tag: Parts of speech, it indicates that How a word functions with its meaning grammatically within sentences.

Step-8: Named entity recognition: process of detecting the named entity such as, person name, movie name, organization name, or location.

Step-9 - chunking - Used to collect individual piece of information into bigger pieces of sentence.

Why NLP is difficult:-

Because ambiguity and uncertainty exist in system

(accent)

grammar mistake
by people

There are three kind of ambiguity -

① Lexical Ambiguity - exist in the presence of two or more possible meaning within single word.

Many a is looking for a match

② Syntactic Ambiguity - exist in the presence of two or more possible meaning within the sentence

I saw the girl the binocular .

③ Referential Referential Ambiguity - exists when you are referring to something referring

Using pronouns.

Kiran went to Sunita, She said I am Hungry.

NLP API -

- ① IBM Watson
- ② Chatbot API
- ③ Speech to Text API
- ④ Sentiment Analysis API
- ⑤ Translation API (Opinion Mining) by Systran
- ⑥ Text analysis API by ~~Ayuan~~ AYUAN
- ⑦ Cloud NLP API
- ⑧ Google Cloud Natural Language API



- NLP Libraries:
- ① Scikit Learn → to build ML Models for NLP
 - ② Natural language Tool Kit → Complete tool kit
 - ③ Pattern → Web mining modules and ML
 - ④ TextBob → easy interface to learn NLP, Sentiment pos tagging.
 - ⑤ Querpy → transforms NL questions into queries in a dbms query language
 - ⑥ Spacy → open source NLP library; used for data extraction, analysis, sentiment analysis.
 - ⑦ Gensim → large dataset and processes data stream

Statistical NLP

Statistical Natural Language Processing) (i)

↳ statistical information extraction from the large corpus of the concerned language can aid in disambiguation



⇒ Corpus ⇒ Collection of written text

A hand-drawn oval shape containing the text "History Data of Google translators".

⇒ Subcorpus ⇒ Some corpora may contain text on particular domain of study or different

$\left\{ \begin{array}{l} \rightarrow \text{Literature} \\ \rightarrow \text{Law} \\ \rightarrow \text{Novels} \end{array} \right\}$

⇒ Parallel Corpus: Some corpora contains a collection of texts which have been translated into one or other several languages.

Counting the elements in Corpus?

N-Grams =

Example -

why wrong - A very wealthy -

Not always get the correct answer

$$\max(P(X | \text{wealthy}))$$

Smoothing: probability -

Spell checker: A spell checker is one of the basic tool required for language processing.

→ Word processing

→ character or text recognition

→ Speech recognition

Example - 326 -

(ii)

Henry sat on the box ①

Henry sat on the box ②,

Spelling Errors:

Three causes of errors -

Insertion

Deletion

Substitution

Three kinds of errors are →

④ Typographic

⑤ Orthographic Errors - a lot of people came ^{two the} _{skool}

③ Phonetics

Why NLP is Hard:

New-1

- ① Ambiguity: Word sense - Bank - River, finance
POS : - chair - Noun, verb

Syntactic Structure → I can see a man

Multiple - I made her duck
→ One morning I shot an elephant ^{with telescope} in pajamas

- ② Scale



- ③ Sparsity - Sparse data due to zip's law
→ The frequency of different words
in large text corpus

- ④ Variation - We train a part of speech tagger
↳ on wall street journal.
↳ on social media -

- ⑤ Expressivity
↳ like such he asked for 40 last names so
he can add me on fb.
→ Not only one form have
different meaning but the same meaning
can be expressed with different form
She gave the book to Tom
↳ She gave Tom the book
Please be quite → shut up.

⑥ Unmodelled Variables

⑦ Unknown representations -

Fields with connection to NLP -

- ① Machine learning
- ② Linguistics
- ③ Cognitive Science
- ④ Information theory
- ⑤ Logic
- ⑥ Data Science
- ⑦ Political Sci
- ⑧ Phylogeny
- ⑨ Economics
- ⑩ Education

Phonetic and Phonology - The study of language sound ^{New-2}

Ecology - The study of language convention for punctuation

Morphology - the study of meaningful component

Syntax - Study of structural relationship

Lexical Semantics - The study of word meaning

Compositional Semantics - the study of meaning of sentence

Pragmatic - The study of the use of language to accomplish goals

Discourse conventions - The study of dialogue

Human Computer Dialogue System

New-3

