

Introduction to Natural Language Processing

Definition:

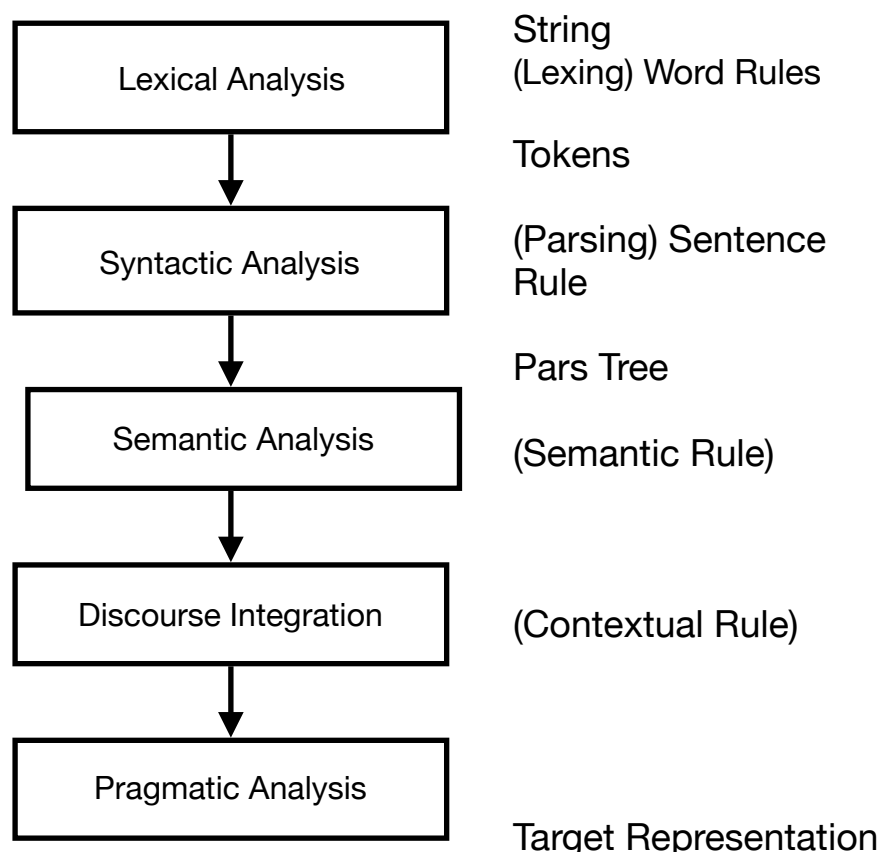
Natural Language Processing is the scientific discipline of computer science and artificial intelligence dedicated to make natural language understandable to machines. And also Interaction between machine and human through Natural-Languages, and specially how to program machines to process enormous natural language data. The input and output of an NLP system can be Speech and written text.

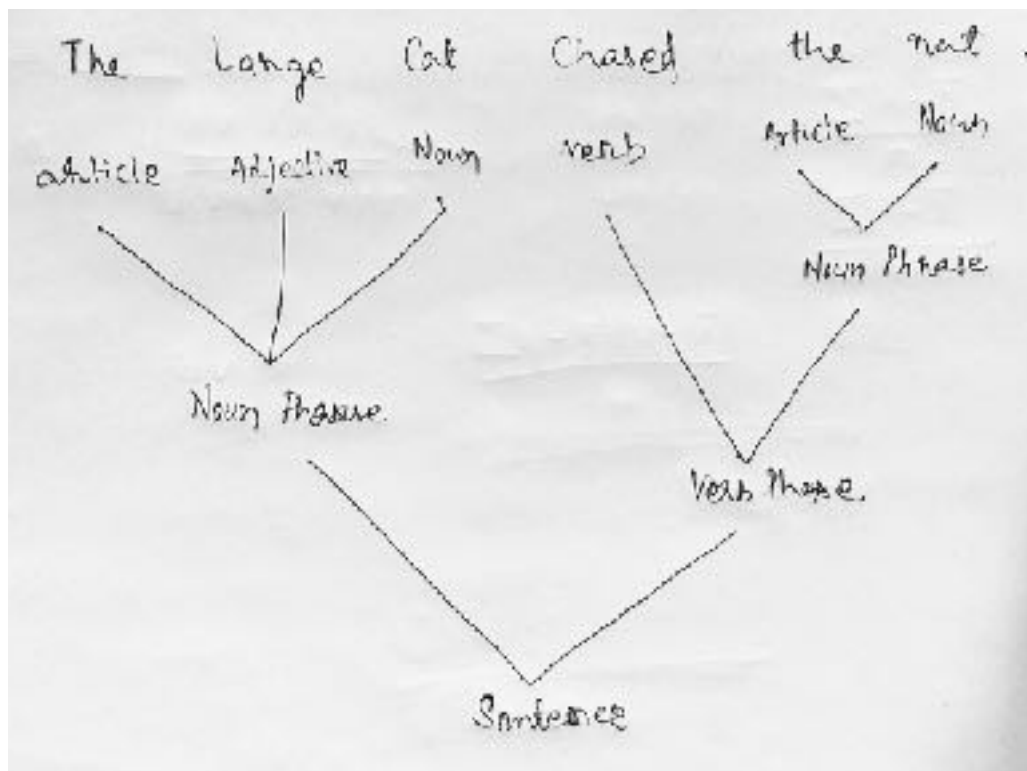
NLP Problems:

Machine translation, Named entity identification, Optical Character Recognition, Question Answering, Relationship Extraction, Sentimental Analysis, Text Summarisation etc.

The tasks involved in NLP:

Lexical Analysis or Morphological Processing: Identifying and analysing the structure of words. Lexical analysis is dividing the whole chunk of txt into paragraphs, sentences and words.
Lexicon- the collection of words and phrases in a language.





Grammar

Sentence \rightarrow NounPhrase, VerbPhrase

VerbPhrase \rightarrow Verb, NounPhrase

NounPhrase \rightarrow Article, Noun

NounPhrase \rightarrow Article, Adjective, Noun

Lexicon

Word	Category
Cat	Noun
Chase	Verb
Large	Adjective
The	Article
Rat	Noun

Syntactic Analysis: Analysis of words in the sentence for grammar and arranging words in a manner that shows the relationship among the words.

Eg: The sentence such as “The Market goes to Soumya” is rejected by English syntactic analyser.

Semantic Analysis: It find the exact meaning or dictionary meaning from the text. The text is checked for meaningfulness. It is done by mapping syntactic structures and objects in the task domain.

Eg: The sentence such as “hot ice-cream” is rejected by semantic analyser

Discourse Integration: This deals with the meaning of immediately succeeding sentence, because mostly the meaning of sentence depends upon the meaning of the sentence just before it.

Pragmatic Analysis: It involves deriving those aspects of language which require real world knowledge.

Eg: *I always return to home, when sunsets.* Need to identify that *I always return to home in evening.*

NLP with Python:

Python is a powerful language and it supports large number of libraries for Machine Learning, data science and natural language processing too.

- A lot of great NLP related Libraries are
NLTK; TextBlob; spaCy; scipy; pandas;
- Parsing Libraries like
BeautifulSoup; Scrapy;..

Major Steps:

- Get Data
- Exploratory Data Analysis
- Visualisation
- Text Pre-Processing
 - Remove Stop Words
 - Remove punctuations
 - Tokenisation
 - Stemming
- Vectorization
- Input to the MODEL