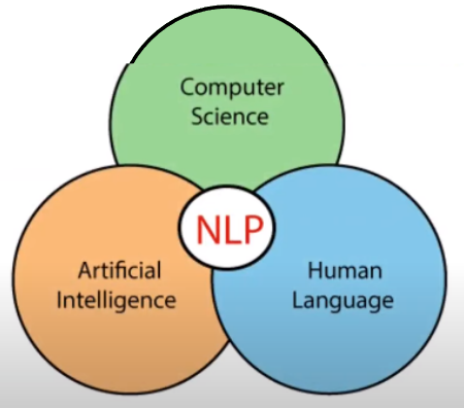
Introduction to NLP - NLP LECTURE 1
Introduction

Author: Harshavardhan

INSTRUCTOR: Nitish Sir, CampusX

Introduction to NLP

Natural language processing is a subfield of linguistics, computer science, and artificial intelligence concerned with the interactions between computers and human language, in particular how to program computers to process and analyze large amounts of natural language data.



Need for NLP

In neuropsychology, linguistics, and the philosophy of language, a natural language or ordinary language is any language that has evolved naturally in humans through use and repetition without conscious planning or premeditation. Natural languages can take different forms, such as speech or signing. They are distinguished from constructed and formal languages such as those used to program computers or to study logic.

Real World Applications

- Contextual advertisements
- Email clients - spam filtering, smart reply
- Social media - removing adult content, opinion mining
- Search engines
- Chatbots

Common NLP tasks

- Text/Document classification
- Sentiment Analysis
- Information Retrieval
- Parts of Speech tagging
- Language detection and machine translation
- Conversational agents
- Knowledge graph and Q&A systems
- Text Summarization
- Topic modeling
- Text Generation
- Spell checking and grammar checking
- Text parsing
- Speech to text

Approaches to NLP

1. Heuristic approaches

A heuristic approach is a problem solving method that uses rules of thumb or educated guesses to find solutions, often sacrificing optimality for efficiency. It is based on intuition and practical experience rather than rigorous mathematical calculations.

Ex:

- rule based- number of positive and negative words
- Regular expressions
- Wordnet
- Openmind common sense

2. Machine learning approach

The Big Advantage: Algorithm figures out the rules.

ML workflow: Text vectorization => ML Algorithm => Output => Evaluation

Algorithms used:

- Naive Bayes
- Logistic Regression
- SVM
- LDA (Topic modeling)
- Hidden markov models

3. Deep Learning approach

The Big Advantage:

- Preserves the Sequential information unlike ML
- Feature generation is automatic

Architectures:

- RNN
- LSTM
- GRU (Text Generation)
- CNN (Text Classification)
- Transformers (BERT)
- Autoencoders (encoder - decoder)

Challenges in NLP

- Ambiguity

  Ex: I saw the boy on the beach with my binoculars.

  I have never tasted a cake quite like that one before!

- Contextual words

  Ex: I ran to store because we ran out of milk

- Colloquialisms and slang

  Ex: Piece of cake, Pulling your leg

- Synonyms
- Irony, Sarcasm and tonal difference

  Ex: That's just what I needed today

- Spelling Errors
- Creativity

  Ex: Poems, dialogue, scripts

- Diversity