

Basics of Natural Language Processing (NLP)

1. Introduction to NLP

Natural Language Processing (NLP) is a branch of Artificial Intelligence (AI) that enables computers to understand, analyze, interpret, and generate human language. Human language is complex, unstructured, and context-dependent, which makes NLP challenging and important.

2. Why NLP is Important

NLP allows machines to interact with humans in a natural way. It is used in applications such as chatbots, search engines, machine translation, sentiment analysis, spam detection, and voice assistants like Alexa and Siri.

3. Levels of NLP

- Lexical Analysis: Deals with words and tokens.
- Syntactic Analysis: Deals with grammar and sentence structure.
- Semantic Analysis: Deals with meaning of words and sentences.
- Pragmatic Analysis: Deals with real-world context and intent.

4. NLP Pipeline

The NLP pipeline is a step-by-step process to convert raw text into useful information.

4.1 Text Preprocessing

- Tokenization: Splitting text into words or sentences.
- Lowercasing: Converting text to lowercase.
- Stopword Removal: Removing common words like 'is', 'the', 'and'.
- Stemming: Reducing words to root form (e.g., playing → play).
- Lemmatization: Converting words to meaningful base form (e.g., better → good).

4.2 Feature Extraction

Since machines understand numbers, text must be converted into numerical form.

- Bag of Words (BoW)
- TF-IDF (Term Frequency – Inverse Document Frequency)
- Word Embeddings (Word2Vec, GloVe, FastText)

5. Common NLP Tasks

- Text Classification (spam detection, sentiment analysis)
- Named Entity Recognition (NER)
- Part-of-Speech (POS) Tagging
- Machine Translation
- Text Summarization

6. Traditional NLP vs Modern NLP

Traditional NLP uses rule-based and statistical methods such as BoW and TF-IDF. Modern NLP uses deep learning models like RNNs, LSTMs, and Transformers such as BERT and GPT, which understand context better.

7. Challenges in NLP

- Ambiguity in language
- Sarcasm and irony
- Multiple languages and dialects
- Informal text and slang

8. Conclusion

Natural Language Processing plays a vital role in modern AI systems. Understanding NLP basics is essential for building intelligent applications that interact with humans using natural language.