

I Introduction to NLP.

* Origin and history of NLP.

- NLP refers to **AI method** of communicating with an intelligent system using a natural language such as English.
- Processing of natural language is required when you want an intelligent system like robot to perform as per our instructions.
- natural language understanding (NLU) is a subtopic of NLP in AI that deals with machine reading comprehension.
- Goal of NLP is to design and build software that will **analyze, understand and generate** languages that humans use naturally.
- current approaches to NLP are based on machine learning, AI that examines & uses patterns in data to improve a program's own understanding.

* Level of NLP.

- morphology: analysis of individual words that consist of morphemes, smallest unit of grammar
- syntax: it is concerned with rules. It includes legal formulation of sentences to check valid structures.
- semantic: meaning check is carried out after structure

- discourse integration: generally, meaning of current sentence is dependent on the one that is prior to it. This deals with analysis of discourse structure.
- pragmatic: deals with mapping of interpreted meaning to actual meaning.
- prosody: analysis phase that handles rhythm, most difficult
- phonology: involves analysis of different kinds of sounds that are combined.

* Stages in NLP.

① Lexical analysis.

- scan source as stream of characters and convert to meaningful lexemes.
- studies patterns of formation of words

② Syntactic analysis.

- used to check grammar, word arrangements & shows the relationship among words.
- words → phrases → clauses → sentences

③ Semantic analysis.

- concerned with meaning representation
- focuses on the literal meaning, studies meaning of words independent of sentence's context
- may involve ambiguity.

④ Discourse integration.

- depends upon sentences that precede it and also follow it, connects sentences
- focuses on inter-sentential connections, studies how previous sentence can alter the interpretation of next sentence.

⑤ Pragmatic knowledge.

- last phase of NLP, helps one to discover the actual, intended meaning by applying a set of rules.

* Knowledge types.

① Factual - terminologies, glossaries, details

② Conceptual - understanding of principles of a domain

③ Procedural - knowledge of how to perform a specific skill or task, knowledge related to methods, procedures or operation of equipment.

④ meta cognitive - learner's knowledge about learning.

* Ambiguity

- ability of being understood in more than one way.

① Lexical ambiguity: ambiguity of a single word is called "-".
e.g. silver as noun or adjective.

② Syntactic ambiguity: occurs when a sentence is parsed in different ways.

- ③ Semantic ambiguity: occurs when meaning of words themselves can be misinterpreted; happens when a sentence contains ambiguous ~~as~~ word or phrase
- ④ Anaphora ambiguity: arises due to the use of ~~an~~ anaphoric entities in discourse.
- ⑤ Pragmatic ambiguity: when context of a phrase gives it multiple interpretations

* Challenges of NLP

- Language differences: different languages have different vocabularies and different phrasing, need to retrain NLP models.
- Training data: NLP is about analysing language, abilities of an NLP system depend on the training data provided to it.
- Development time
- Phrasing ambiguities: system must be able to seek out context that can help understand the phrasing.
- Misspellings
- Innate biases: system would be biased because of data set or programmes.

- Words and phrases with multiple meanings
- False positives and uncertainty.

✶ Applications of NLP.

① Translation

- complicated, not just word-to-word method.
- must translate without changing the meaning of sentence
- Google translate once used phrase-based machine translation (PBMT), now uses Google neural MT (GNMT)

② Speech recognition

- ability to recognise speech & interpret words & phrases from spoken language.

③ Sentiment analysis

- uses NLP to interpret and analyse emotions in subjective data like news, etc
- identify +ve, -ve & neutral points.

④ Chatbots

- employ pattern recognition with heuristic responses, used to hold conversation with humans.

⑤ Question-answer systems

⑥ automatic text summarization

⑦ Market intelligence

⑧ automatic text classification