

SPPU-BE-COMP-CONTENT – KSKA Git

Q1: Explain NLP. Why it is Hard?

ANS.

• NLP stands for 'Natural Language Processing.'

• Natural Language Processing is a branch of Artificial Intelligence (AI) that focuses on enabling computers to understand, interpret, process and generate Human Language in a Meaningful Way.

• It combines concepts from linguistics, Computer Science, and ML to work with text data.

• NLP used in application such as Language translation, text summarization, etc.

— Why NLP is Hard?

(1.) Ambiguity

— Words and sentences can have Multiple Meaning.

(2.) Context Dependence.

— Meaning often depends on previous sentences or real world knowledge.

(3.) Variable Language.

• The same idea can be expressed in many ways using different words, grammar or styles.

(4.) Implicit Knowledge

• Humans use common sense and world knowledge, which is difficult for machines to Model.

Q2: Differentiate between Programming Language's and Natural Language

ANS.

→

SPPU-BE-COMP-CONTENT – KSKA Git

NO.	PROGRAMMING LANGUAGE:-	NO.	NATURAL LANGUAGE:-
1.	These are Formal Languages used to write instructions for computers.	1.	These are Languages used by humans for communication.
2.	They have strict Syntax and Grammar Rules.	2.	Grammar rules are often flexible and often ambiguous.
3.	No Ambiguity Allowed.	3.	Ambiguity is common.
4.	Meaning is exact and precise.	4.	Meaning depends on context.
5.	Used to communicate instructions to machine.	5.	Communication is between Humans.
6.	Errors cause program Failure.	6.	Errors are Tolerable.

Q3. > Are Natural Languages Regular? Explain in Detail.

ANS.

• No., Natural Languages are not Regular.

• A Regular Language is one that can be completely described using regex and recognized by infinite Automation.

- These Languages have limited memory and cannot handle the complex sentence structure.

Now,

- Natural Languages are not Regular because:-

(1.) Long Distance Dependencies.

• Natural Languages require matching elements over long distances, which Finite Automata can't handle.

SPPU-BE-COMP-CONTENT - KSKA Git

2. Nested and Recursive structures,

- Natural Languages allow recursion's

3. Agreement Constraints:

- Subject verb Agreement needs memory of Number and person.

Q4. > Explain the Following Terms:-



(i) Finite Automata for NLP.

- A Finite Automata is a Mathematical Model with a finite number of states used for recognizing the pattern in the text.

- Role of NLP:-

- Used for Lexical Analysis.
- Word Recognition.
- Token Validation.
- Simple Pattern Matching.

(ii) Stages of NLP.

- The various stages of NLP are:-

① Lexical Analysis.

Ex: Breaking the input (text/sentence) into words or Atomic tokens.

② Syntactic Analysis (Parsing)

- Checking grammatical structures.

③ Semantic Analysis.

- Determining meaning of sentences.

④ Disclosure Integration.

SPPU-BE-COMP-CONTENT - KSKA Git

- Understanding meaning across sentences.

⑤ Pragmatic Analysis:-

- Understanding context and real world meaning.

(iii) Challenges and issues in NLP:-

→ 1. Ambiguity

- Lexical, syntactic, and semantic Ambiguity.

2. Context Understanding.

- Meaning changes with context.

3. Language Diversity.

- Different Languages, dialects and styles.

4. Informal Language.

- Slang, Abbreviations, spelling mistakes.

Q5. → What is the concept of tokenization, stemming, lemmatization and POS Tagging. Explain all the terms with suitable examples.

ANS.

1. Tokenization.

Tokenization is the process of breaking text into smaller unit called Tokens (word / phrase / letter)

For Eg:-

→ Text = " My name is Rohan " (Input)

Output:- After tokenization.

['My', 'name', 'is', 'Rohan']

↑ ↑ ↑ ↑
(Tokens)

SPPU-BE-COMP-CONTENT – KSKA Git

2. Stemming:-

- stemming is the process of reducing words to their root form by removing suffixes.
- The word resulting from stemming may or may not be valid.
- It is fast but sometimes inaccurate.

→ For Example:-

- ① Fisher → Fish
- ② studies, studying → study.
- ③ Argue, Arguing, argued → argu

3. Lemmatization:-

- Lemmatization reduces words to their base or dictionary form i.e. (lemma) by considering grammar and meaning.
- The Output is always a 'Valid Word'.

→ For Example:-

- ① Dogs → Dog
- ② Better → Good.
- ③ Runs → Run
- ④ Was → Be.
- ⑤ Mice → Mouse.

4. POS Tagging:- (Parts-of-Speech Tagging)

- POS Tagging assigns grammatical labels (such as nouns, verbs, adjective) ~~such as~~ to each word in a sentence.
- It helps in understanding sentence structure and meaning.

→ For Example:-

I	like	to	read	Books.
↓	↓	↓	↓	↓
(pronoun)	(verb)	(to)	(verb)	(Noun)