

NLP Answer Key – IAT 1

1A	How does NLP bridge the gap between human communication and machine understanding? • NLP develops computational models to help machines process and understand human language. • It automates language processing and improves understanding of human communication. • NLP involves both language interpretation and generation, including speech processing. • Linguistics and psycho-linguistics contribute to NLP, aiding in language structure and comprehension. • By integrating computational linguistics, NLP enables machines to understand and generate human-like text and speech.
1B	Discuss any 5 major challenges in Natural Language Processing (NLP). 1. Problems of representation and interpretation 2. The inability to capture all the required knowledge is another source of difficulty 3. Greatest source of difficulty in natural language is identifying its semantics 4. Idioms, metaphor and ellipses add more complexity to identify the meaning of the written text 5. Quantifier scoping 6. Ambiguity 7. Various sources of ambiguities in natural languages
2A	What are the different levels of language processing, and what type of knowledge does each involve? 1. Lexical Analysis-Morphological knowledge 2. Syntactic Analysis-Syntactic Knowledge: 3. Semantic Analysis: pragmatic information 4. Discourse Analysis: Knowledge of the world
2B	Consider the following Corpus: 1. John likes coffee 2. John drinks tea 3. Tea is good Find the probability of the sentence "John likes tea" using the bigram model. $P(\text{like} \text{John}) = 0.5$ $P(\text{tea} \text{likes}) = 0$ the probability of the sentence "John likes tea" using the bigram model is "0"
3A	Identify the Karaka roles in the following sentence based on Karaka theory: "<i>Ravi wrote a letter with a pen in the library yesterday.</i>" 1. Ravi - subject – Karta 2. a letter – object – Karma 3. with a pen - instrument – Karana 4. in the library – location – Adhikarana
3B	Write the challenges in processing Indian Languages. 1. Unlike English, Indian Scripts have a non linear structure. 2. Unlike English, Indian languages have Subject Object Verb (SOV) as a default sentence structure. 3. Indian languages have a free word order i.e, words can be moved freely within a sentence without changing the meaning of the sentence.

	- Spelling Standardization is more subtle in Hindi than in English. That is, it is delicate in meaning or intent and difficult to understand. - Indian Languages have a rich set of morphological variants. - Indian Languages makes extensive and productive use of complex predicates (CPs). - Indian Languages uses post-position (Karakas) case markers instead of preposition. - Indian Languages uses verb complexes consisting of sequences of verbs.
4A	Write a Regular Expression to validate a vehicle registration number that follows this pattern: - Starts with two upper case letters (eg. KA, TN) - Followed by two digits (eg. 25,07) - Followed by two upper cases(optional) - Ends with four digits (eg. 1234) Test your answer with two examples (valid and invalid). $^[A-Z]{2}\backslash d{2}[A-Z]{0,2}\backslash d{4}\$ Explanation of the Regex: $^$ → Start of the string. $[A-Z]{2}$ → First two uppercase letters (State code like KA, TN). $\backslash d{2}$ → Two digits (e.g., 25, 07). $[A-Z]{0,2}$ → Optional two uppercase letters (e.g., AB or can be missing). $\backslash d{4}$ → Four digits at the end (e.g., 1234). $\$ → End of the string.
4B	Draw state diagram for the following languages: $L = (aa)^*(bb)^*$ $L = (a+b)^*c$
	Compute the minimum edit distance between ' <i>paecflu</i> ' and ' <i>peaceful</i> '.

5A	play + V + PP). 4. Key components: Lexicon, morphotactics, and orthographic rules. 5. Two-level model (Koskenniemi, 1983) uses Finite State Transducers (FSTs).
6A	Explain stochastic tagging. • Stochastic have data-driven approaches in which frequency-based information is automatically derived from corpus and used to tag words. • Stochastic taggers disambiguate words based on the probability that a word occurs with a particular tag. • The simplest scheme is to assign the most frequent tag to each word. • Example for Stochastic tagger was CLAWS (Constituent Likelihood Automatic Word-tagging System), which is equivalent to TAGGIT. • Hidden Markov Model (HMM) is the standard Stochastic tagger.
	Give a possible parse tree for the sentence: <i>"The old book was found in the library."</i>

6B

