BIRLA INSTITUTE OF TECHNOLOGY, MESRA, RANCHI
(END SEMESTER EXAMINATION)

CLASS: MTECH **SEMESTER : II**
BRANCH: CSE **SESSION : SP/2025**

SUBJECT: CS633 NATURAL LANGUAGE PROCESSING

TIME: 3 Hours **FULL MARKS: 50**

INSTRUCTIONS:
1. The question paper contains 5 questions each of 10 marks and total 50 marks.
2. Attempt all questions.
3. The missing data, if any, may be assumed suitably.
***4*. Before attempting the question paper, be sure that you have got the correct question paper.**
5. Tables/Data hand book/Graph paper etc. to be supplied to the candidates in the examination hall.

			CO	BL
Q.1(a)	Explain the different phases of Natural Language Processing (NLP) and discuss why ambiguity is a key challenge in NLP.	[5]	1	2
Q.1(b)	Define Tokenization, Normalization, Lemmatization, and Stemming with suitable examples. How do these contribute to NLP preprocessing?	[5]	1	2
Q.2(a)	What is an N-gram language model? Explain with an example how a bigram model is constructed and used.	[5]	2	3
Q.2(b)	Discuss the issue of unknown words in N-gram models. How does smoothing help in such cases? Briefly describe any one basic smoothing technique.	[5]	2	3
Q.3(a)	Define Term Frequency-Inverse Document Frequency (TF-IDF) and Pointwise Mutual Information (PMI). How are they useful in semantic analysis?	[5]	3	2
Q.3(b)	Explain the working of the Word2Vec model. How can it capture the semantic similarity between words?	[5]	3	3
Q.4(a)	Describe the basic concepts of Part-of-Speech (POS) tagging. What are word classes and tagsets in this context?	[5]	4	2
Q.4(b)	How does the Hidden Markov Model (HMM) work in POS tagging? Illustrate using a simple example	[5]	4	3
Q.5(a)	Differentiate between top-down and bottom-up parsing strategies. Give an example to illustrate both.	[5]	5	4
Q.5(b)	What is the CKY parsing algorithm? Describe its working with reference to context-free grammar.	[5]	5	3

::::::01/05/2025::::::E