

Programme Name & Branch: MIC/MID

Course Name & code: Natural Language Processing and Computational Linguistics [CSI4001]

Class Number (s): VL2024250502130/ VL2024250502122

Faculty Name (s): Dr. Biji C.L./Dr. Sharmila Banu

Exam Duration: 90 Min.

Max Marks

Q. No.	Question	Max Marks	CO
1.	<p>Given the following mini corpus,</p> <p>Document 1: Chatbots can significantly enhance customer service by providing 24/7 support! Can you believe it?</p> <p>Document 2: Incorporating machine learning algorithms, chatbots can learn and adapt to users' preferences. Let's go!!</p> <p>(i) Explain the sentence segmentation algorithm with the mini corpus given (5 Marks)</p> <p>(ii) Compute the type token ratio (3 Marks)</p> <p>(iii) List function words and content words (2 Marks)</p>	10	CO1
2.	<p>Identify the type of ambiguity with possible interpretations and ways to handle ambiguity for the following sentences.</p> <p>(i) The bank can loan money.</p> <p>(ii) She said he would help her.</p> <p>(iii) The bear is running.</p> <p>(iv) The burglar threatened the student with a knife.</p>	10	CO1
3.	<p>Calculate the minimum edit distance between the following pairs of strings using dynamic programming: ISRO and IRON. Consider the insertion cost=1, deletion cost=1 and substitution cost=2. List any four applications of minimum edit distance</p>	10	CO2
4.	<p>Analyze and understand the given test statement using the text preprocessing techniques.</p> <p>Have fun learning NLP.</p> <p>Compare and contrast between stemming and lemmatization with respect to text preprocessing with suitable examples</p>	10	CO2
5.	<p>Consider the given mini training corpus</p> <p><s> the cat sleeps on the ground</s></p> <p><s>the cat runs and sleeps</s></p> <p><s>the cat sits on mat</s></p> <p>Test data:</p> <p><s> the cat sleeps on mat </s></p> <p>Compute the probabilities of test sentence based on the unigram and bigram language model</p>	10	CO3