VIT
Vellore Institute of Technology

School of Computer Science and Engineering
Fall Weekend Intra Semester 2024-25
Continuous Assessment Test – 1

SLOT:C1+TC1

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: 50**

Q. No.	Question	Max Marks	CO	BL
1.	Given the following mini corpus, **Document 1:** Chatbots can significantly enhance customer service by providing 24/7 support! Can you believe it? **Document 2:** Incorporating machine learning algorithms, chatbots can learn and adapt to users' preferences. Let's go!! (i) Explain the sentence segmentation algorithm with the mini corpus given (5 Marks) (ii) Compute the type token ratio (3 Marks) (iii) List function words and content words (2 Marks)	10	CO1	BL2
2.	Identity the type of ambiguity with possible interpretations and ways to handle ambiguity for the following sentences. (i) The bank can loan money. – Pragmatic (ii) She said he would help her. – Semantic (iii) The bear is running. – Lexical (iv) The burglar threatened the student with a knife. – Pragmatic	10	CO1	BL4
3.	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	10	CO2	BL3
4.	Analyze and understand the given test statement using the text preprocessing techniques. **Have fun learning NLP.** Compare and contrast between stemming and lemmatization with respect to test preprocessing with suitable examples	10	CO2	BL1
5.	Consider the given mini training corpus <s> the cat sleeps on the ground</s> <s>the cat runs and sleeps</s> <s>the cat sits on mat</s> **Test data:** <s> the cat sleeps on mat </s> Compute the probabilities of test sentence based on the unigram and bigram language model	10	CO3	BL3