



VIT

Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

REG.NO.:

SLOT: C1+TC1

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING CONTINUOUS ASSESSMENT TEST - I WINTER SEMESTER 2025-2026

Programme Name & Branch : B.Tech Computer Science and Engineering
Course Code and Course Name : BCSE409L Natural Language Processing
Faculty Name(s) : Dr. SARAVANAKUMAR K, Dr. KANNADASAN R, Dr. RAJESHKANNAN R,
 Dr. SATHYARAJ R, Dr. JABANJALIN HILDA J, Dr. ARIVOLI A
Class Number(s) : VL2025260502034, 2035, 2032, 2033, 2039, 2270
Date of Examination : 29-01-2026
Exam Duration : 90 minutes Maximum Marks: 50

General instruction(s):

Answer All Questions

- M - Max mark; CO - Course Outcome; BL - Blooms Taxonomy Level (1 - Remember, 2 - Understand, 3 - Apply, 4 - Analyse, 5 - Evaluate, 6 - Create)
- Course Outcomes
 CO1-Understand the fundamental concepts of Natural Language Processing.
 CO2-Develop useful systems for language processing and related tasks involving text processing and demonstrate text-based processing of natural language with respect to morphology.
 CO3-Check the syntactic and semantic correctness of natural language

Q. No	Question	Module	Marks	CO	BL
1.	A healthcare organization wants to use NLP to automatically generate short summaries from clinical reports to assist doctors in quick diagnosis. For example, from a medical note: <i>"The patient exhibits elevated blood glucose levels and persistent fatigue, indicating possible type-2 diabetes."</i> Explain in detail how the different stages of NLP will process this input to generate a concise medical summary.	1	10	CO 1	BL2
2.	Identify the type(s) of ambiguity present in the following sentences and justify your answers with appropriate explanations. a) The teacher discussed the exam with the students in the hall. b) Anita saw the girl with the binoculars. c) Rahul informed Suresh that he had won the competition. d) Can you pass the salt? e) After Neha spoke to Riya, she said she would resign from the project.	1	10	CO 1	BL2
3.	a) Design a Finite State Transducer (FST) to model the K-insertion rule for English verbs ending with -c when forming the past tense. i. Describe the states and input-output transitions of the FST. (2 marks) ii. Explain how the FST inserts 'k' only for verbs ending with -c and maps other verbs to their regular past-tense forms without insertion. (3 marks) (Examples: mimic → mimicked, panic → panicked, traffic → trafficked, play → played) b) . Given the following mini corpus, Document 1: Dr. Sharma joined the AI lab in 2021. He published 3.5 papers per year on avg.	2	5	CO 2	BL3



VIT

Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

REG.NO.:

SLOT: C1+TC1

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING
CONTINUOUS ASSESSMENT TEST - I
WINTER SEMESTER 2025-2026

	<p>Document 2: The system achieved an accuracy of 98.7%! Can it be improved further? Yes, it can. Explain the decision-tree-based sentence segmentation algorithm and show how it segments the above corpus. (5 marks)</p>	2	5	CO 2	BL3												
4.	<p>Given the following transition and emission probabilities, use the Viterbi algorithm to find the most probable tag sequence for the sentence: "time flies like arrows"</p> <table border="1" data-bbox="323 908 1476 1268"> <thead> <tr> <th>Emission Probabilities</th> <th>Transition Probabilities</th> </tr> </thead> <tbody> <tr> <td>$P(\text{"time"} N)=0.30$</td> <td>$P(V N)=0.5$</td> </tr> <tr> <td>$P(\text{"flies"} V)=0.20$</td> <td>$P(P V)=0.4$</td> </tr> <tr> <td>$P(\text{"flies"} N)=0.05$</td> <td>$P(N P)=0.8$</td> </tr> <tr> <td>$P(\text{"like"} P)=0.40$</td> <td>$P(N N)=0.2$</td> </tr> <tr> <td>$P(\text{"arrows"} N)=0.60$</td> <td></td> </tr> </tbody> </table> <p>The initial probabilities are: $P(N) = P(V) = P(P) \approx 0.33$ Perform the following operations. (i) Construct the Viterbi trellis (ii) Show calculations at each step (iii) Identify the best tag sequence</p>	Emission Probabilities	Transition Probabilities	$P(\text{"time"} N)=0.30$	$P(V N)=0.5$	$P(\text{"flies"} V)=0.20$	$P(P V)=0.4$	$P(\text{"flies"} N)=0.05$	$P(N P)=0.8$	$P(\text{"like"} P)=0.40$	$P(N N)=0.2$	$P(\text{"arrows"} N)=0.60$		2	10	CO 2	BL3
Emission Probabilities	Transition Probabilities																
$P(\text{"time"} N)=0.30$	$P(V N)=0.5$																
$P(\text{"flies"} V)=0.20$	$P(P V)=0.4$																
$P(\text{"flies"} N)=0.05$	$P(N P)=0.8$																
$P(\text{"like"} P)=0.40$	$P(N N)=0.2$																
$P(\text{"arrows"} N)=0.60$																	
5.	<p>A news aggregation company is building a rule-based NLP system to extract key phrases (Noun Phrases and Verb Phrases) from short news headlines. Because headlines are short, linear, and mostly follow fixed POS patterns, the development team decides to use a Finite State Transducer (FST) for shallow parsing instead of a full syntactic parser. The system receives POS-tagged input and must output chunk labels. Consider the following headline: "The skilled engineer designed a bridge" The chunking rules followed by the system are:</p> <ul style="list-style-type: none"> $NP \rightarrow (DET) (ADJ) N^*$ $VP \rightarrow V (NP)?$ <p>Perform the following operations:</p> <ol style="list-style-type: none"> Assign each token with the POS tag for the given sentence. Design a Finite State Transducer (FST) for identifying Noun Phrase (NP) chunks. <ul style="list-style-type: none"> Clearly mention the states, transitions, and accepting states. Extend the FST to handle Verb Phrase (VP) chunking. Apply the designed FST to the given POS-tagged sentence and generate the labels for each word and write the final shallow-parsed output by grouping words into chunks. Discuss the advantages and limitations of shallow parsing techniques and explain the applications of shallow parsing in NLP systems. [0+3+2+2+3] 	3	10	CO 3	BL4												