VIT®

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

Final Assessment Test(FAT) - NOV/DEC 2025

Programme	B.Tech.	Semester	Fall Semester 2025-26
Course Code	BCSE409L	Faculty Name	Prof. Krithiga R
Course Title	Natural Language Processing	Slot	D1+TD1
		Class Nbr	CH2025260100534
Time	3 hours	Max. Marks	100

Instructions To Candidates

- Write only your registration number in the designated box on the question paper. Writing anything elsewhere on the question paper will be considered a violation.

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.
CO4: Select a suitable language modelling & Feature Representation to develop real-world applications.
CO5: Develop computational methods for real-world applications using deep learning.

Section - I
Answer all Questions (4 × 10 Marks)

01. Perform Porter's Stemming Algorithm on the highlighted words **(complaining, deliveries, responded, resolved, expressed, satisfaction, recommended, future, purchases)** and justify each step with the corresponding rule(Step 1a–5) that applies in the algorithm. Explain briefly how stemming benefits your feedback analysis system in terms of feature extraction and dimensionality reduction.

    [10] (CO2/K2)

02. Develop a hybrid rule-based + lexicon-driven Finite State Automata (FSA) that generates verb forms for both regular and irregular verbs such as:

Verb	Past	Past Participle	Present Participle
walk	walked	walked	walking
go	went	gone	going
run	ran	run	running

    a) Construct an FSA diagram illustrating how transitions occur between regular and irregular verb generation paths. [5 Marks]
    b) Identify and label the major states in your automaton corresponding to lexical lookup, morphological suffixation, and output formation.[5 Marks]

    [10] (CO2/K2)

03. A Natural Language Processing (NLP) research team is developing a dependency parser to analyze sentence structures for an automated grammar feedback system. To test the parser, they input the sentence:
    "Students submitted the assignment before the deadline."
    (a) As part of the analysis, identify the head word for each dependency relation and represent the dependency structure using an arc diagram. [5 Marks]
    (b) Explain how the dependency parser determines the root of the sentence during parsing. [5 Marks]

    [10] (CO3/K3)

04. You are building an automatic summarization system for news articles. Consider the following original text
    "Global financial markets reacted cautiously after the Federal Reserve announced a potential interest rate hike, as investors assess the impact on inflation and economic growth."
    Generated Summary:
    "Markets reacted cautiously after Fed hints at interest rate increase."
    Reference Summaries:
    - *"Investors are cautious following Fed's potential interest rate hike affecting economic growth."*
    - *"Financial markets show caution due to possible Fed interest rate increase."*

a. Compute **ROUGE-1 and ROUGE-L** scores of the generated summary against each reference summary. [7 Marks]

b. Show precision, recall, and F1. [3 Marks]

[10] (CO5/K3)

Section - II
Answer all Questions (4 × 15 Marks)

05. A biomedical NLP researcher is applying Byte Pair Encoding to tokenize specialized medical terms from patient reports.

The given corpus is:

"antibacterial, antibiotic, antitoxin, autoimmune, antiviral, vaccination"

a) Simulate 5 BPE merge iterations, showing the main merges and resulting subword vocabulary. [6 Marks]

b) Identify whether BPE learns meaningful morphemes like anti, auto, toxin, viral, vaccin. [6 Marks]

c) Compare the resulting BPE segments with rule-based morphological analysis for two words. [3 Marks]

[15] (CO1/K3)

06. You are building a parser for English sentences using a probabilistic Context Free Grammar (CFG). Consider the grammar:

Rule	Probability
S → NP VP	1.0
NP → Det N	0.5
NP → NP PP	0.5
VP → V NP	0.6
VP → VP PP	0.4
PP → P NP	1.0
Det → the	0.7
Det → a	0.3
N → man	0.4
N → telescope	0.6
V → saw	1.0
P → with	1.0

Sentence: "the man saw the man with the telescope"

a) Draw 2 parse trees using the CFG rules. [7 Marks]

b) Compute the probability of each parse using the PCFG. [5 Marks]

c) Identify the most probable parse. [3 Marks]

[15] (CO3/K4)

07. Let as assume that, You have 4 news headlines:
- *"Stock market hits record high today"*
- *"Tech stocks soar after new product launch"*
- *"Global economy shows signs of recovery"*
- *"Market analysts predict downturn in tech sector"*

a) Construct a Term Frequency - Inverse Document Frequency (TF–IDF) matrix for the 4 headlines. [9 Marks]

b) List the top 2–3 terms per headline by TF–IDF weight and discuss how TF–IDF can help group similar headlines. [6 Marks]

[15] (CO3/K3)

08. You are designing a bigram model for a personal assistant that handles emails and scheduling. The corpus includes:

"Schedule a meeting with the marketing team for next Wednesday at 3 PM"

"Send an email to the client summarizing the last project update"

"Reschedule my dentist appointment to Thursday afternoon"

"Remind me to submit the budget report before Friday"

a) List all unique bigrams and their counts. [6 Marks]

b)Compute Maximum Likelihood Estimation (MLE) bigram probabilities for all observed bigrams. [9 Marks]

[15] (CO4/K4)

BL-Bloom's Taxonomy Levels - (K1-Remembering,K2-Understanding,K3-Applying,K4-Analysing,K5-Evaluating,K6-Creating)

⟺⟺⟺