Z/K/TY

Reg. No: 22BLE0415

VIT
Vellore Institute of Technology

# Final Assessment Test – November 2025

Course: BCSE409L - Natural Language Processing

Class NBR(s): 1746/1755/1760/1773/2654 Slot: D2+TD2

Time: Three Hours Max. Marks: 100

- KEEPING MOBILE PHONE/ANY ELECTRONIC GADGETS, EVEN IN 'OFF' POSITION IS TREATED AS EXAM MALPRACTICE
- DON'T WRITE ANYTHING ON THE QUESTION PAPER

| COs | CO Statements |
|---|---|
| CO1 | Understand the fundamental concept 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 sematic correctness of natural language. |
| CO4 | Select a suitable language modeling and feature representation to develop real world applications. |
| CO5 | Develop computational methods for real world applications using deep learning. |

**BL – Blooms Taxonomy Level (1 – Remember, 2 – Understand, 3 – Apply, 4 – Analyse, 5 – Evaluate, 6 – Create)**

**Answer ALL Questions**
**(10 X 10 = 100 Marks)**

1. Assume you are the CTO of a healthcare analytics company. You are handed over a large corpus of patient feedback and hospital reviews collected from multiple sources (online portals, surveys, and mobile apps). Explain precisely what kind of NLP stage-wise outcomes and insights can be determined from this data. Highlight how these insights can support hospital management, patient experience teams, and clinical operations. **CO1 BL2**

2. Design an FSA for Adjectives in English demonstrating Derivational Morphology. Illustrate at least with three examples. Provide a supporting lexicon and show the transitions for words in the lexicon. Identify the exceptions, if any that could not be addressed by the FSA. **CO2 BL2**

3. Calculate the minimum edit distance between the following pair of strings using dynamic programming: "flaw" and "lawn". Show the dynamic programming matrix and compute the final edit distance. Also, list any four practical applications of minimum edit distance in computational linguistics or data science. **CO2 BL3**

4. Consider the following CFG that describes a simple language: **CO3 BL3**

Grammar:

S → NP VP

NP → PRP | Det Adj N | N | PRP NP

VP → V NP | V NP PP

PP → P NP

Det → 'a' | 'the'
Adj → 'missed'
N → 'call' | 'morning'
PRP → 'She' | 'me'
V → 'gave' P → 'in'

Derive the parse trees for the ambiguous sentence *"She gave me a missed call in the morning."*

Explain the different interpretations and briefly discuss methods to resolve this ambiguity in practical NLP systems.

5. How does Maximum Entropy perform Part-of-Speech tagging? Present the Mathematical Model, explain the notations, working and its advantages over the HMM tagging? CO3 BL2

6. Discuss any two methods to compute semantic similarity between words or concepts in NLP. Explain how hierarchical knowledge bases and corpus statistics can influence similarity outcomes. CO3 BL3

7. Explain the process of Semantic Role Labeling (SRL) and its relevance in Natural Language Processing using the following sentence: CO4 BL2

*Anna sent her friend a postcard from Paris last summer*

a) Identify the semantic roles of the relevant constituents in the sentence.
b) Describe the steps involved in performing SRL for this sentence, including the role of syntactic parsing.

8. Consider the following training corpus of student feedback: CO4 BL3

- The professor explained the concept of entropy clearly.
- The instructor gave a detailed lecture on transformers.
- The tutor helped us understand the chatbot architecture.

Test Corpus

i. The professor explained the chatbot.
ii. The tutor gave a lecture on entropy.

a) Analyse the limitation in solving the probability of sentence (i) and (ii)
b) Resolve the limitation and compute the probability of sentence (i) and (ii)

9.a) Given the sentence: *"Birds can fly in clear skies."* CO4 BL4

i. Using a sliding window of size 2, list all the possible context-target pairs for CBOW training.
ii. Describe the CBOW architecture (layers, parameters, and their dimensions) used to process one training pair.
iii. Take the context ["Birds", "can", "in", "skies"] and target "fly". Using a vocabulary size of 6 and an embedding dimension of 3, explain the calculation of the hidden vector, output scores, softmax probabilities, and cross-entropy loss (show all steps, assuming parameter matrices are provided)

**OR**

| | | | | |
|---|---|---|---|---|
| 9.b) | Design a Named Entity Recognition (NER) system that adapts to user-specific domain preferences over time, such as recognizing custom entity types (e.g., product codes, internal team names). Explore techniques for personalization in NER models. Discuss the challenges of maintaining generalization across domains, handling ambiguous or nested entities, and balancing user-specific adaptations with overall system robustness and scalability. | CO4 | BL4 |
| 10.a) | Explain a Case study on Information Extraction and its application in Text Summarization. Discuss the pipeline, extracted entities and how they influence the summarization outcome. | CO4 | BL3 |

**OR**

| | | | |
|---|---|---|---|
| 10.b) | Explain a case study on sentiment analysis using deep learning. Highlight how feature representation, model interpretability and emotion granularity affect the overall accuracy and ethical deployment of the system. | CO4 | BL3 |

⇔⇔⇔ **Z/K/TY** ⇔⇔⇔