Y/K/TY

Reg. No:

VIT
Vellore Institute of Technology

Final Assessment Test - November 2025

Course: **BCSE409L - Natural Language Processing**

Class NBR(s): **1741/1744/1751/1766/1771**

Slot: **D1+TD1**

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 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 problems.
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. A digital library wants to use NLP to summarize research papers for students. For an instance, from the abstract: "This paper presents a machine learning model for predicting exam **performance using** attendance and **assignment** data." Elaborate how the stages of NLP will handle this input in detail. **CO1** **BL3**

2. Given the following root words, generate **two** inflectional and two derivational forms of each. Explain the changes in detail. Root words: **play, Strong.** **CO2** **BL2**

3. Determine the PoS tag for the given test data using VITERBI algorithm. [Training data and Test data is given below] **CO2** **BL4**

Training Corpus

<s>	Book	a	car	</s>
	verb	Det	Noun	
<s>	Read	a	book	</s>
	verb	Det	Noun	
<s>	book	is	Heavy	</s>
	Noun	verb	Adj	

Test data

words	Book	a	book
POS tags	Noun	Det	Noun
	verb		verb

4. a) Identify the type of semantic relationship for each of the following word pairs. [5] CO3 BL3

i. Hot – Cold
ii. Car – Wheel
iii. Rose – Flower
iv. Buy – Purchase
v. Bank (financial institution) – Bank (river side)

b) Explain in detail how Word Sense Disambiguation (WSD) is performed using the Lesk Algorithm. [5]

5. Justify the need for word embeddings in modern NLP systems with a well labelled diagram, explain the architecture and training mechanism of FastText. Critically analyze its advantages and limitations compared to word2Vec and contextual embedding models. CO3 BL3

6. An e-commerce company wants to automatically classify product reviews like "The phone is light but the battery drains fast" into positive, negative, or neutral categories. Their older TF-IDF, Word2Vec model struggles because it treats words independently and fails to understand context. Identify a more suitable model for this task and explain its working mechanism in detail. CO4 BL4

7. Given the training corpus, Determine the most probable sentence among 4 candidate sentences using bigram model. CO4 BL4

Training corpus

<s> students play football</s>
<s> students play chess</s>
<s> Teachers teach students</s>

Candidate sentences:

<s>students play football</s>
<s>students play chess </s>
<s>teachers teach students </s>
<s>teachers play football </s>

8. A news agency wants to use Named Entity Recognition (NER) to automatically tag entities in articles. Elucidate how NER would process the below sentence and why it is useful for news applications. CO5 BL4

Sentence:

"Prime Minister Narendra Modi visited New Delhi on 15th August 2025 to attend Independence Day celebrations at Red Fort."

9.a) Parse the sentence "The boy eats rice with curd" using transition based dependency parsing. Show the parsing steps and draw the final dependency structure. CO3 BL3

OR

9.b) i. Perform regular-expression based chunking on the sentence: "The little boy quickly ate an apple." by using the following rules [5] CO3 BL3

NP → DT JJ* NN
VP → V NP
AdvP → RB

ii. Is the sentence *'Visiting relatives can be annoying"* ambiguous? Justify your answer with possible interpretations. [5]

10.a) A media company has a system that takes long political articles and produces a short, 3-4 sentence version for quick reading. For example, from a 1,000 word article about government policies, the system creates a concise summary that captures the main points without losing meaning. Identify what type of NLP application is suitable for the given scenario. Elaborate how it works and discuss why it is useful for readers. CO5 BL4

OR

10.b) A university is building a chatbot to answer student queries such as: When is the last date to submit assignments?" "Who is the head of the Computer Science department?" CO5 BL4

The chatbot must read the university's frequently asked questions(FAQ) and handbook and provide direct answers in natural language. Identify which model is most suitable for this Question answering task. Justify your answer and explain how the model processes a question and produces the answer in detail.

⇔⇔⇔ Y/K/TY ⇔⇔⇔