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

Final Assessment Test(FAT) - Nov/Dec 2024

Programme	B.Tech.	Semester	Fall Semester 2024-25
Course Code	BCSE409L	Faculty Name	Prof. Lakshmi Harika Palivela
Course Title	Natural Language Processing	Slot	B2+TB2
		Class Nbr	CH2024250101621
Time	3 hours	Max. Marks	100

General Instructions

- Write only Register Number in the Question Paper where space is provided (right-side at the top) & do not write any other details.

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 (8 × 10 Marks)

*M - Marks

Q.No	Question	*M	CO	BL
01.	Imagine you are developing a text correction algorithm for a document editing application. The application needs to suggest corrections for words that have been mistyped. In this scenario, the user has typed the word *"ELEPHANT"*, but the intended correction should be *"RELEVANT"*. (a) Determine the minimum number of individual changes required to convert *"ELEPHANT"* into *"RELEVANT"*. (8 Marks) (b) Analyze the time and space complexity of your solution. (2 Marks)	10	1,2	4
02.	Consider the given corpus to calculate the probabilities for the given tag sequence using the Viterbi algorithm model and find the most probable tag sequence for the sentence, "The park is a book." Corpus: \<S> Book a car \</S> \<S> Park the car \</S> \<S> The book is in the car \</S> \<S> The car is in a park \</S>	10	2	3

03. Elaborate on the working of a two-step morphological parser in a Finite State Transducer. Also, construct the FST diagram for the following words: 10   3   2

a) Radiantly (2 Marks)
b) Merging (2 Marks)
c) Computational (2 Marks)
d) Luminously (2 Marks)
e) Coulrophobia (2 Marks)

04. Consider the following grammar G which is written in Chomsky's normal form. Find out the probability of generating the sentence using the Probabilistic Context-Free Grammars (PCFG) algorithm.   10   3   5

S = "Papa ate the caviar with a spoon".
S → NP VP (1)
NP → Det N (0.5)
NP → NP PP (0.25)
VP → V NP (0.5)
VP → VP PP (0.5)
PP → P NP (1)
NP → Papa (0.25)
N → caviar (0.5) | spoon (0.5)
V → spoon (0.5) | ate (0.5)
P → with (1)
Det → the (0.5) | a (0.5)

05. Determine if the following sentence is ambiguous. If it is, draw both the phrase structure and dependency parse trees for each possible interpretation of the sentence: "The gunman killed police in Chennai."   10   2,3   3

06. (a) Provide an example sentence and identify the semantic roles of each argument with proper justification. (6 Marks)   10   3,4   2

(b) Compare two different Semantic Role Labelling (SRL) frameworks/systems (e.g., FrameNet vs. PropBank) and discuss with respect to key features such as: focus, roles, sensitivity and scalability. (4 Marks)

07. Consider the below Corpus: 10   4   3
"The cat sits on the mat."
"The dog barks at the cat."
"The dog chases the cat."
"The cat sits on the wall."

a) Compute the Bi-gram probability for all the sentences. (4 Marks)
b) Apply Laplace smoothing for "The cat sits on the mat". (3 Marks)
c) Calculate perplexity for "The dog chases the cat". (3 Marks)

08. As an NLP expert, analyze how AI-assisted chatbots utilize dialogue state architecture to facilitate interactions between users and AI-powered assistance. Illustrate the various components of dialogue state architecture with relevant example.   10   1,5   2

Section - II
Answer all Questions (1 × 20 Marks)

Question

Q.No	Question	*M	CO	BL
09.	(a) Suppose you have the following two sentences of 4-dimensional word vectors, which are X_1 and X_2 respectively: X_1: "AI transforms future possibilities" the vector values are (0.4; 0.2; 0.3; 0.8), X_2: "Machines learn from data" the vector values are (0:5; 0:3; 0:5; 0:9). Evaluate cosine similarity between X_1 and X_2? Are the words in X_1 and X_2 similar or dissimilar? (5 Marks)	20	4	3

(b) For the corpus given below, train a TFIDF model to find the context-based word embedding and explain the steps involved in it. (15 Marks)

"AI shapes modern technology advancements. Data drives AI insights. Automation transforms various industries."

BL-Bloom's Taxonomy Levels - (1.Remembering, 2.Understanding, 3.Applying, 4.Analysing, 5.Evaluating, 6.Creating)

⇔⇔⇔