BG/D/TZ

Reg. No: 0439

VIT®
Vellore Institute of Technology

Final Assessment Test – April 2026

Course: **BCSE409L - Natural Language Processing**

Class NBR(s): **2032/2033/2034/2035/2039/2270** Slot: **C1+TC1**

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. A service robot is used in a hospital to assist doctors and staff. The robot can understand spoken commands from staff. If the staff gives the instruction: **CO1 BL2**

"Bring the medicine to room five."

The robot must process this instruction, understand the words in the command, analyze the sentence structure, determine the meaning of the instruction, and identify the correct action to perform.

Explain how a Natural Language Processing system helps the robot understand and execute the given command. Describe the stages involved in interpreting the instruction and converting it into an appropriate action performed by the robot. Illustrate your explanation using the given example.

2. a) Compare and contrast the Inflectional and derivation morphology. **[5] CO2 BL3**

b) Discuss the following in brief with simple example, **[5]**

i. Stemming

ii. Lemmatization

3. Consider a smart home system wants to predict whether the house is in "High Energy" or "Low Energy" mode based on observable appliance usage. The actual energy mode is hidden, but the system observes the following activities: Using AC, Cooking, or Watching TV. CO3 B

Hidden States: High Energy, Low Energy
Observations: AC, Cooking, TV
The system has learned the following probabilities.

Initial Probabilities
P(High Energy) = 0.6
P(Low Energy) = 0.4

Transition Probabilities

From / To	High Energy	Low Energy
High Energy	0.7	0.3
Low Energy	0.4	0.6

Emission Probabilities

Activity	High Energy	Low Energy
AC	0.6	0.1
Cooking	0.3	0.4
TV	0.1	0.5

Suppose the observed sequence of activities is:

AC → Cooking → TV

Compute the probability of each state at every step and determine the most likely sequence of energy modes using the Viterbi algorithm. Draw the optimal state path.

4. Explain Dependency Parsing in Natural Language Processing. Discuss the concept of dependency grammar and dependency trees. Explain the different types of dependency parsing with a suitable example sentence and describe how head-dependent relationships represent syntactic structure in a sentence. CO3 BL4

5. A system analyzes social media posts to find related discussions. **CO3 BL3**

Document	Post
D1	climate change affects environment
D2	environment protection reduces climate change
D3	football team wins final match

Vocabulary = {climate, change, environment, protection, football, match}
Construct the Term Frequency (TF) matrix for the documents, compute the Document Frequency (DF) for each term, and calculate the IDF values using log(N/DF). Then determine the TF-IDF vectors for D1 and D2, compute the Cosine Similarity between them, and interpret whether the two posts discuss the same topic.

6. i. Discuss the following word sense relations with suitable examples: **[6] CO3 BL4**
   a) Synonymy and Antonymy
   b) Hypernymy and Hyponymy
   c) Polysemy and Homonymy

   ii. Construct a hierarchical tree showing the hypernym–hyponym relationship for the word "Food". **[4]**

7. A company is developing a NLP based customer support chatbot to predict the next word in a user message using a bigram language model. Since some word combinations do not appear in the training data, Laplace Smoothing is applied to avoid the zero-probability problem. **CO4 BL4**
The chatbot was trained with the following conversation dataset:

No	Sentence
S1	customer needs help
S2	customer needs support
S3	user needs help
S4	customer wants support
S5	user wants help
S6	customer needs assistance
S7	user needs support
S8	customer requests help

Vocabulary (**V**) = {customer, user, needs, wants, requests, help, support, assistance}

a) Construct the unigram and bigram frequency table for the word "needs". Determine the total occurrence count of the word "needs." Using Laplace smoothing, calculate the following probabilities: [6]

P(help | needs)

P(support | needs)

P(assistance | needs)

P(customer | needs)

b) Using the calculated bigram probabilities, compute the probability of the sentence: [4]

Test sentence:

customer needs help

8. A space research organization is developing an AI system to analyze satellite communication and mission reports. The system uses an NLP language model trained on space exploration documents to identify patterns in mission communication and scientific data. CO4 BL4

Training Corpus:

<s> satellites collect space observation data </s>
<s> satellites transmit space communication signals </s>
<s> astronauts explore deep space missions </s>
<s> satellites support space research missions </s>
<s> astronauts collect space experiment data </s>
<s> satellites monitor earth observation data </s>
<s> astronauts conduct space research experiments </s>
<s> satellites provide space weather data </s>
<s> satellites collect earth observation images </s>
<s> astronauts analyze space experiment results </s>

Test Sentence:

satellites collect space experiment data

Calculate the trigram probabilities of the test sentence and determine the perplexity of the sentence. Show all the trigram probability calculations and compute the final perplexity value.

9.a) "A large airline company collects customer feedback from surveys, emails, and online travel websites. The feedback includes opinions about flight services, staff behavior, food quality, and punctuality. The management wants to analyze these comments to understand customer satisfaction and improve service quality. Therefore, the company plans to implement a sentiment analysis system using Natural Language Processing to automatically identify positive and negative opinions in customer feedback and generate insights for decision-making". CO5 BL4

Based on this scenario, explain how sentiment analysis can help the airline company understand customer satisfaction, describe the process of building a sentiment analysis system for analyzing customer feedback, and discuss the evaluation methods that can be used to measure the performance of the sentiment analysis model.

OR

9.b) "A global e-commerce company operates in multiple countries and serves customers who speak different languages. The platform contains thousands of product descriptions, customer reviews, and support messages that need to be available in many languages so that customers can easily understand the information. Currently, the company relies on manual translation, which is time-consuming, costly, and difficult to manage as the amount of content continues to grow. To address this challenge, the company plans to develop an automatic machine translation system using Natural Language Processing that can translate multilingual text while preserving the original meaning and context". **CO5 BL4**

Based on this scenario, explain how an NLP-based machine translation system can be developed by applying concepts such as text preprocessing, tokenization, language modeling, and neural machine translation to produce accurate translations while maintaining contextual meaning.

10.a) "In the famous online learning platforms provide thousands of lecture transcripts, course materials, and discussion forum posts for students. Many students find it difficult to go through long lecture transcripts to identify the key concepts required for exam preparation. To assist students, the platform plans to implement an NLP based automatic summarization tool that can generate meaningful summaries of lecture transcripts and discussion threads. The system should find the main ideas, keep the meaning correct, and present the information clearly and briefly. Developers are considering modern transformer-based models along with traditional NLP methods to improve the quality of summaries". **CO5 BL2**

Examine the above scenario and explain in a simple way how an NLP-based text summarization system can summarize lecture transcripts with a clear step-by-step procedure. Discuss the role of preprocessing, extractive and abstractive summarization techniques, and suggest suitable evaluation methods to measure the quality of generated summaries.

OR

10.b) "A government organization maintains a large collection of public policy documents, legal guidelines, and citizen service manuals. Citizens often struggle to find answers to their questions because the documents are long and complex. To make information more accessible, the organization plans to build an NLP based Question Answering system that allows users to ask questions in simple language and receive precise answers from the official documents. The system should identify the most relevant sections of the documents and extract accurate answers. Developers are considering the use of modern deep learning techniques and large language models to improve the accuracy and efficiency of the system". CO5 BL2

Examine the above scenario and discuss how a Question Answering system can help citizens access information more easily. Discuss the stages involved in processing the user query, retrieving relevant information, and generating the final answer with proper justification.

⇔⇔⇔ BG/D/TZ ⇔⇔⇔