

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

NOIDA INSTITUTE OF ENGINEERING AND TECHNOLOGY, GREATER NOIDA
(An Autonomous Institute Affiliated to AKTU, Lucknow)

B.Tech

SEM: VII - THEORY EXAMINATION (2025 - 2026)

Subject: Natural Language Processing

Time: 3 Hours

Max. Marks: 100

General Instructions:

IMP: Verify that you have received the question paper with the correct course, code, branch etc.

1. This Question paper comprises of **three Sections -A, B, & C**. It consists of Multiple Choice Questions (MCQ's) & Subjective type questions.

2. Maximum marks for each question are indicated on right -hand side of each question.

3. Illustrate your answers with neat sketches wherever necessary.

4. Assume suitable data if necessary.

5. Preferably, write the answers in sequential order.

6. No sheet should be left blank. Any written material after a blank sheet will not be evaluated/checked.

SECTION-A

20

1. Attempt all parts:-

1-a. Which of the following is NOT a challenge in NLP?(CO1,K1)

1

(a) Ambiguity

(b) Perfect accuracy

(c) Data quality

(d) Scalability

1-b. In NLP, what does the acronym BERT stand for?(CO1,K2)

1

(a) Basic English Reading Tool

(b) Bidirectional Encoder Representations from Transformers

(c) Best Entity Recognition Technique

(d) Binary Encoding and Recognition Tool

1-c. What is the role of part-of-speech tagging in NLP? (CO2,K1)

1

(a) a) Assigning grammatical categories to words

(b) b) Handling URLs

(c) c) Converting text to lowercase

(d) d) Replacing repeated characters

1-d. In N-gram models, what does a trigram represent? (CO2,K2)

1

(a) a) A sequence of two adjacent words

(b) b) A sequence of N adjacent words or tokens

(c) c) A sequence of three adjacent words or tokens

(d) d) A single word or token

1-e. Which text vectorization model represents documents as vectors in a high-

1

dimensional space?(CO3,K3)

- (a) Bag-of-Words model
 - (b) Vector Space model
 - (c) TF-IDF model
 - (d) Word Mover's Distance
- 1-f. Which of the following is a similarity measure commonly used with vector space models?(CO3,K2) 1
- (a) Jaccard similarity
 - (b) Hamming distance
 - (c) Levenshtein distance
 - (d) Pearson correlation
- 1-g. In the context of topic modeling, what is a "document-topic distribution"?(CO4,K2) 1
- (a) A list of authors for each document
 - (b) The probability distribution of topics for each document in a corpus
 - (c) The number of words in each document
 - (d) The length of the documents
- 1-h. In neural machine translation, what type of architecture is commonly used for modeling translation tasks?(CO4,K2) 1
- (a) Convolutional Neural Network (CNN)
 - (b) Recurrent Neural Network (RNN)
 - (c) Support Vector Machine (SVM)
 - (d) Transformer
- 1-i. How does LSTM differ from traditional RNNs?(CO5,K1) 1
- (a) LSTM has no recurrent connections
 - (b) LSTM can handle long-term dependencies better
 - (c) LSTM is designed for image processing
 - (d) LSTM lacks sequential processing capability
- 1-j. Which of the following is an architecture specifically designed for attention mechanisms in NLP?(CO5,K2) 1
- (a) RNN
 - (b) CNN
 - (c) Transformer
 - (d) LSTM

2. Attempt all parts:-

- 2.a. Define the term "text mining."(CO1,K1) 2
- 2.b. Why is it important to have access to diverse corpora in linguistic research?(CO2,K2) 2
- 2.c. Discuss one real-world application where Word2Vec is commonly used.(CO3,K1) 2
- 2.d. In the context of text classification, define the term "neutral sentiment."(CO4,K2) 2
- 2.e. State two uses of BERT in NLP tasks.(CO5,K3) 2

SECTION-B

30

3. Attempt all parts:-	
3.a. Answer any <u>one</u> of the following:-	
3.a.(i) Explain the role of NLP in bridging the gap between human language and computational analysis.(CO1,K3)	6
3.a.(ii) Discuss the historical evolution of NLP and how it has transformed over the years.(CO1,K2)	6
3.b. Answer any one of the following:-	
3.b.(i) Discuss how linguistic analysis is vital in advancing natural language understanding and generation, with a focus on its impact on AI applications, such as chatbots and virtual assistants.(CO2,K3)	6
3.b.(ii) Describe the role of regular expressions in text preprocessing.(CO2,K2)	6
3.c. Answer any one of the following:-	
3.c.(i) Define TF-IDF and elaborate on its significance in addressing the limitations of the Bag-of-Words model.(CO3,K2)	6
3.c.(ii) Explain the Bag-of-Words model and how it represents documents in text vectorization. Discuss its advantages and limitations.(CO3,K3)	6
3.d. Answer any one of the following:-	
3.d.(i) Contrast rule-based and statistical approaches in machine translation. Discuss the advantages and disadvantages of each approach, providing examples.(CO4,K3)	6
3.d.(ii) Explain the role of statistical models in machine translation.(CO4,K2)	6
3.e. Answer any one of the following:-	
3.e.(i) Examine the application of sequence-to-sequence models in text summarization tasks. Discuss the challenges associated with summarizing long documents and propose solutions.(CO5,K3)	6
3.e.(ii) Differentiate between unidirectional and bidirectional RNNs.(CO5,K2)	6
<u>SECTION-C</u>	50
4. Answer any <u>one</u> of the following:-	
4-a. Explain the concept of "overfitting" in NLP models and its impact on model generalization and performance.(CO1,K2)	10
4-b. Discuss the purpose and algorithms behind stemming, and provide examples of how it simplifies text normalization.(CO1,K3)	10
5. Answer any <u>one</u> of the following:-	
5-a. Explain the differences between semantics and syntax in linguistic analysis and provide examples illustrating their distinctions.(CO2,K3)	10
5-b. Discuss the impact of emojis in text data and their role in sentiment analysis. How can NLP systems effectively process and interpret emojis?(CO2,K3)	10
6. Answer any <u>one</u> of the following:-	
6-a. Discuss strategies for handling rare words in vector space models. Evaluate the impact of rare words on the overall quality of document representation.(CO3,K3)	10
6-b. Discuss the challenges and best practices in building language models, including the selection of training data and model evaluation techniques.(CO3,K3)	10
7. Answer any <u>one</u> of the following:-	

- 7-a. Discuss an incident where a high-profile organization faced a phishing attack that went undetected by its spam detection system. Analyze the consequences and propose measures to prevent future incidents.(CO4,K4) 10
- 7-b. Explain the concept of sequential data and discuss why traditional neural networks fail to handle it effectively.(CO4,K2) 10
8. Answer any one of the following:-
- 8-a. Discuss the application of reinforcement learning in sequence models, particularly in NLP tasks. Provide an example where reinforcement learning enhances the performance of a sequence model.(CO5,K3) 10
- 8-b. Compare RNN, LSTM, and Transformer architectures. Explain with suitable examples where each is preferable.(CO5,K2) 10

REG_JULY_DEC_2025