Printed Page:- 04 Subject Code:- ACSAI0712 /ACSAIH0712

Roll. No:

NOIDA INSTITUTE OF ENGINEERING AND TECHNOLOGY, GREATER NOIDA

(An Autonomous Institute Affiliated to AKTU, Lucknow)

B.Tech

SEM: VII - THEORY EXAMINATION (2024 - 2025)

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. What is the purpose of tokenization in NLP?(CO1, K2) 1

(a) To make words uppercase

(b) To split text into words or tokens

(c) To add punctuation to text

(d) To remove stop words

1-b. What is stemming in NLP? (CO1, K1) 1

(a) A way to make text bold

(b) A technique to reduce words to their base or root form

(c) A method for adding synonyms to text

(d) A way to add more context to words in a sentence

1-c. What is the primary purpose of converting text to lowercase in data preprocessing? (CO2, K1) 1

(a) a) To handle URLs

(b) b) To ensure case-insensitive text processing

(c) c) To handle HTML tags

(d) d) To remove stop words

1-d. What is the purpose of removing HTML tags from text data in NLP? (CO2, K2) 1

(a) a) To extract plain text content

(b) b) To handle email IDs

(c) c) To handle URLs

(d) d) To convert text to lowercase

1-e. What is the primary purpose of text vectorization in NLP?(CO3, K2) 1

(a) Visualization of textual data

(b) Representation of text as numerical features

(c) Creation of word clouds

(d) Parsing and tokenization of text

1-f. What does TF-IDF stand for?(CO3,K1) 1

(a) Term Frequency-Inverse Document Frequency

(b) Token Frequency-Inverse Document Factor

(c) Text Frequency-Inverse Data Flow

(d) Term Formulation-Incremental Document Filter

1-g. What is the main purpose of spam detection in text classification?(CO4,K2) 1

(a) Identify malicious software

(b) Categorize messages as spam or not spam

(c) Summarize long texts

(d) Translate text into another language

1-h. Topic modeling is primarily used for:(CO4,K2) 1

(a) Categorizing text into predefined classes

(b) Detecting spam emails

(c) Identifying the main themes in a collection of texts

(d) Generating chatbot responses

1-i. What is the primary purpose of sequence models in NLP?(CO5,K2) 1

(a) Image classification

(b) Handling sequential data

(c) Audio processing

(d) Text summarization

1-j. Expand LSTM in the context of sequence models. (CO5,K1) 1

(a) Long Short-Term Memory

(b) Linguistic Sequence Transfer Model

(c) Limited Short-Term Memory

(d) Long Semantic Text Model

2. Attempt all parts:-

2.a. Discuss how lexical ambiguity differ from syntactic ambiguity in NLP.(CO1,K2) 2

2.b. What is the purpose of replacing repeated characters in text data?(CO2,K2) 2

2.c. What is the fundamental idea behind the Bag-of-Words model in text vectorization?(CO3,K1) 2

2.d. What machine learning algorithm is commonly used for text classification tasks, 2

	including sentiment analysis?(CO4,K1)	
2.e.	Explain the concept of transfer learning in the context of advanced NLP techniques.(CO5,K2)	2

SECTION-B 30

3. Answer any five of the following:-

3-a.	Discuss the key difference between stemming and lemmatization?(CO1,K2)	6
3-b.	Describe the role of stop-word removal in NLP, its impact on text analysis, and its relationship with feature selection.(CO1,K2)	6
3-c.	Explain how handling email IDs in text data preprocessing is relevant in NLP, and discuss potential challenges associated with it. (CO2,K2)	6
3-d.	Elaborate on the significance of removing HTML tags from text data in NLP and provide a step-by-step approach to achieve this.(CO2,K3)	6
3.e.	Compare the training methodologies of GloVe and Word2Vec. In what scenarios might GloVe be preferred over Word2Vec, and vice versa?(CO3,K3)	6
3.f.	Explore the challenges associated with rule-based machine translation. How do these challenges differ from those faced in statistical machine translation?(CO4,K3)	6
3.g.	Explain the challenges associated with training recurrent neural networks (RNNs) on long sequences in natural language processing (NLP) tasks. Propose and discuss potential solutions.(CO5,K2)	6

SECTION-C 50

4. Answer any one of the following:-

4-a.	Explore the "curse of dimensionality" in NLP and how it affects data processing and computational efficiency.(CO1,K2)	10
4-b.	Elaborate on the different types of ambiguity in language, including lexical, syntactic, and semantic ambiguity, and their implications for NLP.(CO1,K3)	10

5. Answer any one of the following:-

5-a.	Explain the concept of word embeddings, their role in capturing word semantics, and the difference between traditional vector spaces and word embeddings.(CO2,K3)	10
5-b.	Discuss the role of data preprocessing in NLP and provide examples of specific tasks where data preprocessing is critical.(CO2,K2)	10

6. Answer any one of the following:-

6-a.	Analyze how similarity measures like Cosine Similarity and Word Mover's Distance are influenced by varying document lengths.(CO3,K4)	10
6-b.	Elaborate on the significance of TF-IDF in text vectorization, addressing the limitations of the Bag-of-Words model. Provide a detailed example scenario to demonstrate its application.(CO3,K3)	10

7. Answer any one of the following:-

7-a.	Compare and contrast Latent Dirichlet Allocation (LDA) and Non-Negative	10

	Matrix Factorization (NMF) in topic modeling. Evaluate their performance in uncovering latent topics in a given text corpus(CO4,K4)	
7-b.	Propose a novel evaluation metric for assessing the quality of topics generated by topic modeling algorithms. Justify your choice and compare it with existing metrics.(CO4,K4)	10

8. Answer any one of the following:-

8-a.	Examine the role of transfer learning in NLP, focusing on how pre-trained models can be fine-tuned for specific tasks. Discuss the challenges and benefits of transfer learning and provide examples of successful applications.(CO5,K4)	10
8-b.	Evaluate the impact of deep learning techniques, specifically RNNs and LSTMs, in processing sequential data for language translation tasks. Discuss challenges and advancements in machine translation using these models(CO5,K4)	10