

**INSTITUTE OF AERONAUTICAL ENGINEERING**

(Autonomous)

Dundigal-500043, Hyderabad

B.Tech VI SEMESTER END EXAMINATIONS (REGULAR) - JULY 2023

Regulation: UG-20

**NATURAL LANGUAGE PROCESSING****Time: 3 Hours**

(COMMON TO CSE(AI&amp;ML) | CSE(DS))

**Max Marks: 70****Answer ALL questions in Module I and II****Answer ONE out of two questions in Modules III, IV and V****All Questions Carry Equal Marks****All parts of the question must be answered in one place only****MODULE – I**

1. (a) Discuss the concept of ambiguity in natural language processing (NLP) and its impact on NLP models and algorithms. How can NLP techniques handle or mitigate ambiguity?  
[BL: Understand| CO: 1|Marks: 7]
- (b) Elucidate how does the integration of knowledge in speech and language processing enhance the accuracy and effectiveness of NLP systems? Can you provide examples of specific applications where this integration has been particularly beneficial?  
[BL: Apply| CO: 1|Marks: 7]

**MODULE – II**

2. (a) How can regular expressions be used to extract specific patterns or words from a given text corpus? Provide examples of common applications of regular expressions in text processing tasks.  
[BL: Understand| CO: 2|Marks: 7]
- (b) Evaluate the methods used for extracting terms from tokens in NLP. How reliable and effective are these methods in identifying key concepts or entities in text? Are there any limitations or biases in information extraction techniques that need to be considered?  
[BL: Apply| CO: 2|Marks: 7]

**MODULE – III**

3. (a) Summarize the process of evaluating language models. What metrics are commonly used to assess the performance and quality of a language model? How can perplexity be used to measure the effectiveness of a language model and its relation to entropy?
- (b) Examine the implications of using huge language models. How do the computational requirements and storage limitations of these models impact their accessibility and scalability? What are the potential biases or ethical concerns that can arise when using such massive language models, and how can they be addressed?  
[BL: Apply| CO: 3|Marks: 7]
4. (a) Outline the concepts of inflection and derivational morphology in language processing. How do these linguistic phenomena contribute to the formation and variation of words? Provide examples to illustrate the differences between inflectional and derivational processes.  
[BL: Understand| CO: 4|Marks: 7]
- (b) Critically assess the role of finite-state morphological parsing in language processing. What are the strengths and weaknesses of this approach in capturing complex morphological phenomena? How does the use of finite-state transducers enhance the efficiency and effectiveness of morphological parsing compared to other techniques?  
[BL: Apply| CO: 4|Marks: 7]

## MODULE – IV

5. (a) Enumerate the concept of supervised disambiguation in NLP. How does Bayesian classification contribute to the disambiguation process? Discuss the role of features, priors, and likelihoods in Bayesian classification for disambiguation. [BL: Understand| CO: 5|Marks: 7]
- (b) Critically analyze the methodological preliminaries in NLP. What are the potential biases or limitations that can arise during data preprocessing and feature extraction? How can these biases impact the accuracy and generalization of NLP models, and what strategies can be employed to mitigate them? [BL: Apply| CO: 5|Marks: 7]
6. (a) Describe dictionary-based disambiguation methods in NLP. How do these methods utilize dictionaries and lexical resources to disambiguate word senses? Discuss the advantages and limitations of dictionary-based disambiguation approaches. [BL: Understand| CO: 5|Marks: 7]
- (b) Reflect on the use of thesauri for disambiguation. What are the challenges in maintaining and updating thesauri to capture the evolving nature of language? How do the limitations of thesauri-based disambiguation methods in handling domain-specific or rare words impact their applicability in different NLP tasks? [BL: Apply| CO: 5|Marks: 7]

## MODULE – V

7. (a) How is the probability of properties calculated in an HMM? Discuss the role of emission probabilities and transition probabilities in determining the likelihood of observed sequences in an HMM. [BL: Understand| CO: 6|Marks: 7]
- (b) Evaluate the information sources used in tagging, specifically in the context of Markov model taggers. How effective are HMM-based taggers in capturing the linguistic information needed for accurate part-of-speech (POS) tagging? Discuss the challenges in incorporating different information sources, such as lexical, contextual, or syntactic, and how these challenges affect the performance of the tagger. [BL: Apply| CO: 6|Marks: 7]
8. (a) Describe the Viterbi algorithm and its role in decoding the most likely tag sequence in HMM-based tagging. How does the Viterbi algorithm efficiently compute the most probable sequence given an input sentence and an HMM? Discuss the computational complexities and trade-offs of the Viterbi algorithm. [BL: Understand| CO: 6|Marks: 7]
- (b) Consider the trade-offs between model complexity and computational resources when applying HMMs to tagging tasks. How does the size and complexity of the language model impact the computational requirements and the ability to scale the tagging system? Discuss strategies for optimizing the efficiency and resource utilization of HMM-based tagging systems in large-scale applications. [BL: Apply| CO: 6|Marks: 7]

