Reg. No. : | | | | | | | | | | | | |

Question Paper Code: U7E02

B.E./B.Tech. DEGREE EXAMINATION, NOV 2024

Seventh Semester

Artificial Intelligence and Data Science

21UAD702-NATURAL LANGUAGE PROCESSING

(Regulations 2021)

PART A - (10 x 2 = 20 Marks)

1	What are the major tasks performed in NLP applications?	CO1-U
2	Apply POS Tagging for the following sentence **"We belong to AI&DS department"**	CO2-App
3	Define FST.	CO1-U
4	Provide two examples of how corpora analysis can benefit language model training in NLP applications.	CO2-App
5	What are the functions of Statistical Estimators in NLP?	CO1-U
6	Perform a t-test to compare the average word lengths between two text samples.	CO2-App
7	Apply methodological preliminaries to set up a basic NLP pipeline for text classification. List the key steps involved.	CO2-App
8	Briefly describe an information-theoretic approach to disambiguation in NLP.	CO1-U
9	What is the basic principle behind a Markov model?	CO1-U
10	Explain how probabilities are used in Hidden Markov Models.	CO1-U

PART - B (5 x 16= 80 Marks)

11 (a)	Discuss why ambiguity makes NLP difficult, and propose methods for resolving syntactic and semantic ambiguities in text processing.	CO1-U	(16)

Or

(b)	Explore the concept of cross-entropy in the context of language models. Apply it to evaluate and compare the performance of various models, discussing the implications of your findings for improving model accuracy.	CO1 -U	(16)

12	(a)	Design a method to detect and handle character encoding mismatches in a dataset containing texts from multiple languages. Include considerations for both automatic detection and manual intervention	CO2-App	(16)
		Or		
	(b)	Develop a Finite State Transducer (FST) for a language that exhibits complex inflectional and derivational patterns. Discuss the design considerations, implementation challenges, and potential applications of your FST model. Provide detailed examples to support your discussion.	CO2-App	(16)
13	(a)	Explain the concept of collocations, types, and their role in enhancing the performance of NLP tasks.	CO1-U	(16)
		Or		
	(b)	Explain the principles of Pearson's chi-square test and likelihood ratios, and discuss how they are applied to assess word associations in a corpus.	CO1-U	(16)
14	(a)	Apply the information-theoretic approach to disambiguate a set of ambiguous words. Calculate measures such as entropy and mutual information for each sense and discuss how these measures help in selecting the most appropriate sense.	CO2-App	(16)
		Or		
	(b)	Implement a multi-method disambiguation system combining Bayesian classification, dictionary-based, and translation-based approaches. Apply it to a set of ambiguous words and assess the improvement in sense resolution accuracy compared to using each method individually.	CO2-App	(16)
15	(a)	Define Markov Modeling in NLP with detail explanation and diagram.	CO1-U	(16)
		Or		
	(b)	Describe the methods used for parameter estimation in Hidden Markov Models, such as the Baum-Welch algorithm. Discuss how these methods adjust parameters to maximize the likelihood of the observed data.	CO1-U	(16)