

		[4]				
OR	iii.	Discuss the process of part-of-speech tagging in natural language processing, highlighting its methodologies, applications, and significance in linguistic analysis.	8	2	02	04 03
Q.4	i.	Provide an overview of speech and phonetics, outlining their fundamental principles.	3	2	02	01 04
	ii.	Analyze the role of the Bayesian method in natural language processing, focusing on its theoretical underpinnings and practical implications.	7	3	04	05 03
OR	iii.	Calculate the minimum edit distance between the words "flaw" and "lawn". With (a) Define strings (b) Initialize a matrix (c) Fill the matrix	7	3	02	02 03
Q.5	i.	Define the concept of smoothing in the context of natural language processing.	4	2	02	01 04
	ii.	Elucidate the concept of N-grams in the context of natural language processing, detailing their significance in text analysis, language modeling, and the statistical representation of word sequences within a given corpus.	6	2	01	02 03
OR	iii.	Given the sentence: "The cat sat on the mat." Calculate the bigrams and trigrams from this sentence and determine the frequency of each n-gram.	6	3	02	01 04
Q.6	Attempt any two:					
	i.	Explain sentiment analysis, including its methodologies, applications, and importance in interpreting emotional tone in textual data.	5	2	02	01 04
	ii.	Explain text classification, methodologies, and its significance in NLP.	5	2	01	02 03
	iii.	Define machine translation and its key methodologies used.	5	2	02	01 04

\*\*\*\*\*

Total No. of Questions: 6

Total No. of Printed Pages:4

Enrollment No.....



Faculty of Engineering  
End Sem Examination Dec 2024

CS3EA12 Natural language Processing

Programme: B.Tech.

Branch/Specialisation: CSE All

Duration: 3 Hrs.

Maximum Marks: 60

Note: All questions are compulsory. Internal choices, if any, are indicated. Answers of Q.1 (MCQs) should be written in full instead of only a, b, c or d. Assume suitable data if necessary. Notations and symbols have their usual meaning.

			Marks	BL	PO	CO	PSO
Q.1	i.	Which approach focuses on using large datasets to automatically learn the patterns of language? (a) Rule-based approach (b) Symbolic approach (c) Heuristic approach (d) Statistical approach	1	2	05	05	03
	ii.	What is a key difference between formal and natural languages in terms of ambiguity? (a) Formal languages are more ambiguous than natural languages (b) Natural languages can have multiple interpretations, while formal languages are designed to avoid ambiguity (c) Formal languages do not have defined syntax rules, while natural languages have strict syntax (d) Natural languages are machine-readable, whereas formal languages are not	1	2	02	02	01
	iii.	Which of the following is a common use of finite state transducers in NLP? (a) Performing syntactic parsing of sentences (b) Mapping phonemes to their corresponding words in speech recognition (c) Assigning part-of-speech tags to words in a sentence (d) Generating deep semantic representations of text	1	2	05	05	03

[2]					
iv.	Which method is commonly used to improve accuracy in part-of-speech tagging?	1	2	05	05 03
	(a) Using predefined dictionaries for all possible words				
	(b) Applying backoff techniques when the most likely tag is not found				
	(c) Ignoring ambiguous words to avoid errors				
	(d) Applying only the most frequent tag to each word				
v.	In a Bayesian framework, how is the posterior probability calculated?	1	2	02	05 04
	(a) By considering only the prior probability				
	(b) By multiplying the likelihood and prior probability, then normalizing				
	(c) By using the sum of the likelihood and prior probability				
	(d) By ignoring the likelihood and using a uniform prior				
vi.	Which algorithm is commonly used to identify and correct spelling errors in text?	1	2	02	03 03
	(a) Viterbi Algorithm				
	(b) Levenshtein Distance				
	(c) PageRank				
	(d) Gradient Descent				
vii.	What is a major drawback of using higher-order N-grams for language modeling?	1	2	02	01 04
	(a) Higher computational cost and data sparsity				
	(b) Poor performance with small datasets				
	(c) Inability to capture word dependencies				
	(d) Ignoring the context of words				
viii.	What is the main purpose of using smoothing techniques in language models?	1	2	02	02 03
	(a) To increase the probability of all unseen events				
	(b) To filter out common words from the dataset				
	(c) To reduce the size of the vocabulary				
	(d) To remove noise from the training data				

[3]					
ix.	How can context help in correcting spelling errors in NLP?	1	3	04	01 04
	(a) By suggesting the most frequently occurring word				
	(b) By analyzing the surrounding words to determine the likely correction				
	(c) By choosing a random word from the vocabulary				
	(d) By replacing the misspelled word with its synonym				
x.	Which of the following challenges can affect the accuracy of sentiment analysis?	1	2	06	02 03
	(a) Simple tokenization				
	(b) Handling of negations and sarcasm				
	(c) Uniform word frequency				
	(d) Ignoring part-of-speech tags				
Q.2	i. Explain the concept of ambiguity in Natural Language Processing (NLP).	2	2	02	04 04
	ii. Differentiate formal language from natural language, highlighting their key distinctions and applications.	3	2	01	01 03
	iii. Elucidate the role of knowledge in the domain of speech and language processing, emphasizing its impact on computational models and linguistic analysis.	5	2	04	01 04
OR	iv. Define Natural Language Processing (NLP) and elaborate on its significance in computational linguistics and artificial intelligence, including its role in enabling machines to perform complex tasks.	5	2	05	05 03
Q.3	i. Elaborate on the process of tokenization within the scope of natural language processing.	2	2	02	01 04
	ii. Provide a comprehensive definition of tokenization and text preprocessing in the context of natural language processing, while also outlining their respective advantages and disadvantages.	8	2	01	05 03

**Marking Scheme**  
**CS3EA12 Natural Language Processing**

Q.1	i)	D. Statistical Approach	1
	ii)	B. Natural languages can have multiple interpretations, while formal languages are designed to avoid ambiguity.	1
	iii)	B. Mapping phonemes to their corresponding words in speech recognition	1
	iv)	B. Applying backoff techniques when the most likely tag is not found	1
	v)	B. By multiplying the likelihood and prior probability, then normalizing	1
	vi)	B. Levenshtein Distance	1
	vii)	A. Higher computational cost and data sparsity	1
	viii)	A. To increase the probability of all unseen events	1
	ix)	B. By analyzing the surrounding words to determine the likely correction	1
	x)	B. Handling of negations and sarcasm	1
Q.2	i.	Concept of ambiguity in Natural Language Processing (NLP)	2
	ii.	(i) Differentiate formal language from natural language	1 mark
		(ii) Highlighting their key distinctions	1 mark
		(iii) Applications	1 mark
	iii.	(i) Elucidate the role of knowledge in the domain of speech and language processing.	3 marks
OR		(ii) Emphasizing its impact on computational models and linguistic analysis.	2 marks
	iv.	(i) Define Natural Language Processing (NLP)	2 marks
		(ii) Elaborate on its significance in computational linguistics and artificial intelligence, including its role in enabling machines to perform complex tasks?	3 marks
			5
Q.3	i.	Process of tokenization?	2
	ii.	(i) Provide a comprehensive definition of tokenization and text preprocessing in the context of Natural Language Processing.	4 marks
		(ii) While also outlining their respective advantages and disadvantages.	4 marks
OR	iii.	(i) Discuss the process of part-of-speech tagging in Natural Language Processing	2 marks
		(ii) Highlighting its methodologies	2 marks
		(iii) Applications	2 marks
		(iv) Significance in linguistic analysis.	2 marks
			8

Q.4	i.	(i) Provide an overview of speech and phonetics	2 marks	3
		(ii) Outlining their fundamental principles	1 mark	
	ii.	(i) Analyze the role of the Bayesian method in Natural Language Processing	4 marks	7
		(ii) Focusing on its theoretical underpinnings and practical implications ?	3 marks	
OR	iii.	Calculate the minimum edit distance between the words "flaw" and "lawn". With		7
		a. Define strings.	2 marks	
		b. Initialize a matrix.	2 marks	
		c. Fill the matrix.	3 marks	
Q.5	i.	Define the concept of smoothing in the context of Natural Language Processing.		4
	ii.	(i) Elucidate the concept of N-grams in the context of Natural Language Processing	2 marks	
		(ii) Detailing their significance in text analysis, language modeling	2 marks	6
		(iii) The statistical representation of word sequences within a given corpus.	2 marks	
	OR	iii.		
Q.6		Given the sentence: "The cat sat on the mat."		6
		(i) Calculate the bigrams.	3 marks	
		(ii) Trigrams from this sentence and determine the frequency of each n-gram.	3 marks	5
	i.	(i) Explain sentiment analysis	2 marks	
		(ii) Methodologies, applications	2 marks	
		(iii) Importance in interpreting emotional tone in textual data.	1 mark	5
	ii.	(i) Explain text classification	2 marks	
		(ii) Methodologies	2 marks	
		(iii) Its significance in NLP.	1 mark	5
	iii.	(i) Define machine translation	3 marks	
		(ii) Its key methodologies used	2 marks	

\*\*\*\*\*