

BACHELOR OF TECHNOLOGY (CBCS) (2021-COURSE)
B. Tech. Sem - VI Computer Science & Engineering : SUMMER : 2025
SUBJECT: NATURAL LANGUAGE PROCESSING

Day : Monday
Date : 26/05/2025

S-25597-2025

Time : 02:30 PM-05:30 PM
Max. Marks : 60

N.B :

- 1) All questions are **COMPULSORY**.
- 2) Figures to the right indicate **FULL** marks.
- 3) Draw neat and labeled diagrams **WHEREVER** necessary.
- 4) Assume suitable data if necessary.

Q.1 Discuss the concept of N-Gram (Bi-gram, Tri-gram) Model for sentence: "about five minutes from dinner....."(Predicted data may be college or class) (10)

OR

Q.1 Define entropy and perplexity in the context of natural language processing. Explain how these measures quantify uncertainty and information content in language data. Discuss the relationship between entropy, perplexity, and language modelling, highlighting their importance in evaluating the predictability and complexity of linguistic data (10)

Q.2 Define concept of Regular expression and Patterns in NLP. Also, write Library functions of Regular expression and Patterns in NLP. Provide examples to illustrate the construction and application of regular expressions in language processing tasks. (10)

OR

Q.2 Define Tokenization. Also explain various methods of tokenization. Explain how raw text data is obtained from various sources and formats. Describe the tokenization process, in text data. Evaluate the challenges and techniques involved in tokenizing raw text effectively. (10)

Q.3 How ambiguity in POS tag is resolved using HMM veterbi algorithm. Solve below example.

Training Corpus:

<s>	Martin	Justin	can	watch	Will	<E>
	N	N	M	V	N	
<s>	Spot	will	watch	martin	<E>	
	N	M	V	N		
<s>	will	Justin	spot	Martin	<E>	
	M	N	V	N		
<s>	Martin	will	pat	spot	<E>	
	N	M	V	M		

Test Data:

<s>	Justin	will	spot	Will	<E>
	N	N	V	N	
		M	N	M	

OR

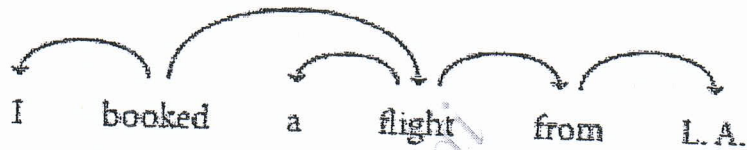
Q.3 Describe the process of acoustic processing in speech recognition, including feature extraction and acoustic modeling. Explain how speech signals are converted into a sequence of feature vectors using techniques. Evaluate the effectiveness of different acoustic processing techniques in capturing the acoustic characteristics of speech signals. (10)

P.T.O.

- Q.4 Use the CKY algorithm for PCFG to find the most probable parse tree. (10)
 Sentence: "The flight includes a meal"
 CFG rules with probabilities
 S → NP VP [0.8]
 NP → Det N [0.3]
 VP → V NP [0.20]
 VP → include [0.05]
 Det → the [0.4]
 Det → a [0.4]
 N → meal [0.001]
 N → flight [0.02]

OR

- Q.4 Describe the concept of Transition based Dependency parser. Solve below (10)
 example using Transition based Dependency parser



- Q.5 Define Word2Vec and Doc2Vec as neural network-based techniques for word and (10)
 document embedding. Explain how Word2Vec learns distributed representations
 of words based on their contextual usage in large text corpora. Discuss how
 Doc2Vec extends Word2Vec to learn document embedding's by considering the
 context of words in documents.

OR

- Q.5 Find the SVD for the matrix $A = \begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{bmatrix}$ (10)

- Q.6 Write Note on OWL, NLTK and Ontologies. (10)

OR

- Q.6 Write notes on Web Technologies and Ontologies (10)

* * * * *