

[This question paper contains 2 printed pages.]

Your Roll No.....

Sr. No. of Question Paper : 7297

L

Unique Paper Code : 2513010024

Name of the Paper : Natural Language Processing

Name of the Course : **B.Sc. (H) Electronics**

Semester : VIII (Under NEP UGCF Mode)

Duration : 3 Hours

Maximum Marks : 90

Instructions for Candidates

1. Write your Roll No. on the top immediately on receipt of this question paper.
2. Question 1 is compulsory.
3. Attempt any 5 questions including question 1.
4. **All** questions carry equal marks.

- Q1 (a) Identify and briefly explain any three major challenges faced in Natural Language Processing. (3)
- (b) Define Morphology and explain its significance in understanding word forms within NLP. (3)
- (c) Differentiate between a Rule-base POS tagger and a Stochastic POS tagger for assigning tags to words. (3)
- (d) Define Stemming and Lemmatization. Give one example of each. (3)
- (e) What are Stop Words? Give 5 examples. (3)
- (f) Differentiate between Bigram and Trigram language models. (3)
- Q2 (a) Trace the origins and fundamental challenges of Natural Language Processing. Why is resolving “ambiguity” (lexical, syntactic, and semantic) considered the most significant challenge in building robust NLP applications? (6)
- (b) Formulate the Trigram Language Model. Assume a training corpus yields the following specific frequency counts: (6)
- $C(\{I, \text{like}\}) = 100$
 - $C(\{I, \text{like}, \text{NLP}\}) = 40$
 - $C(\{I, \text{like}, \text{coding}\}) = 120$
- Based on this data, calculate the trigram probabilities $P(\{\text{NLP}\} | \{I, \text{love}\})$ and $P(\{\text{coding}\} | \{I, \text{love}\})$. Explain the Markov assumption utilized in this calculation.
- (c) Discuss how an empirical comparison of smoothing techniques is conducted and explain why simply assigning a probability of zero to unseen N-grams is detrimental to Applications of Language Modelling. (6)
- Q3 (a) Explain the sequential roles of Tokenization, Named Entity Recognition, and Relation Extraction in converting raw, unstructured text into structured relational data. (6)
- (b) Discuss the concept of Lexical Semantics. How does an understanding of lexical relationships assist in solving the problem of word ambiguity in natural language? (6)

P.T.O.

- (c) Given the sentence “The bank of the river is near the bank where I deposit money”, explain how a Supervised Word Sense Disambiguation model would use context features to distinguish between the two senses of the word “bank”. (6)
- Q4 (a) Elaborate on the Continuous Bag-of-Words (CBOW) model. How does its architecture differ from the skip-gram approach? (6)
- (b) Differentiate between Knowledge-Based WSD (using lexical resources) and Supervised WSD (using annotated corpora) (6)
- (c) Describe the Maximum Entropy model for POS tagging. Discuss how it differs from a standard Markov Model in terms of the features it considers for classification. (6)
- Q5 (a) What is Information Extraction? Explain Tokenization and its role in IE pipeline. (6)
- (b) Define Relation Extraction. Give example of Person-Organization relation. (6)
- (c) What is Inverse Document Frequency (IDF)? Calculate IDF for word "cat" when total documents = 5, documents containing "cat" = 2. (6)
- Q6 (a) Explain Information Retrieval process. How does TF-IDF improve traditional keyword matching? (6)
- (b) List any 4 challenges in Text Summarization. (6)
- (c) What is Sentiment Mining? Explain polarity classification with the dataset of 4 reviews given below: (6)
- 4 Movie Reviews
 Review 1: "Excellent movie, loved every scene!" → POSITIVE
 Review 2: "Very boring, waste of time" → NEGATIVE
 Review 3: "Average film, nothing special" → NEUTRAL
 Review 4: "Great acting but poor story" → NEGATIVE
- The given Keyword List is as follows:
 Positive Keywords: excellent, loved, great, amazing, fantastic
 Negative Keywords: boring, waste, poor, terrible, bad
 Neutral Keywords: average, okay, normal
- Q7 (a) Explain the Need of Machine Translation with 2 real-world applications. Differentiate Direct MT vs Rule-Based MT. (6)
- (b) What is Statistical Machine Translation (SMT)? List 2 advantages and 2 limitations of SMT. (6)
- (c) Explain Encoder-Decoder Architecture in Neural Machine Translation with a diagram/flow. (6)