MAULANA ABUL KALAM AZAD UNIVERSITY OF TECHNOLOGY, WEST BENGAL

Paper Code : BDS502 Natural Language Processing

UPID : 5000014

Time Allotted : 3 Hours *Full Marks :70*

The Figures in the margin indicate full marks.

Candidate are required to give their answers in their own words as far as practicable

Group-A (Very Short Answer Type Question)

1. Answer *any ten* of the following : [1 x 10 = 10]

   (I) Give one disadvantage of n-gram models.

   (II) What is the first step in QA?

   (III) Which task involves extracting meaning from text?

   (IV) What are stop words?

   (V) Which model uses hidden states and observations?

   (VI) What is POS tagging?

   (VII) Give one example of an entity recognized in NER.

   (VIII) Which technique classifies emotions from text?

   (IX) Which MT approach uses neural networks?

   (X) Which IE task extracts subject–predicate–object relations?

   (XI) What type of QA requires factual answers?

   (XII) State one major difference between Natural Language Processing (NLP) and Natural Language Understanding (NLU).

Group-B (Short Answer Type Question)

Answer *any three* of the following : [5 x 3 = 15]

2. What is question classification in QA system? [5]
3. Explain the scope of NLP with respect to syntactic, semantic, and pragmatic levels of language understanding. [5]
4. Differentiate between stemming and lemmatization. [5]
5. What is a Hidden Markov Model (HMM) used for in NLP? [5]
6. Why is stop word removal important in NLP? [5]

Group-C (Long Answer Type Question)

Answer *any three* of the following : [15 x 3 = 45]

7. (a) Explain the difference between **lexicon-based** and **machine learning-based** sentiment classification. [4]

   (b) Illustrate the steps involved in training a machine learning model for sentiment classification using a labeled dataset [4]

   (c) Discuss the role of deep learning models such as LSTM and BERT in sentiment classification. Highlight their advantages over traditional methods. [4]

   (d) Suggest evaluation metrics used for sentiment classification and explain their significance. [3]

8. (a) Explain the concept of Machine Translation (MT). Discuss its importance in natural language processing and real-world applications. [5]

   (b) Compare and contrast Rule-Based Machine Translation (RBMT), Statistical Machine Translation (SMT), and Neural Machine Translation (NMT) in terms of architecture, advantages, and limitations. [7]

   (c) Describe at least two evaluation metrics used for MT systems and explain how they assess translation quality. [3]

9. (a) Define Relation Extraction and explain its significance in building knowledge graphs. [4]

   (b) Describe the main methods for relation extraction: supervised, semi-supervised, and distant supervision. Highlight their advantages and limitations. [6]

(c) Given the sentence:
"Tim Cook is the CEO of Apple Inc., which is headquartered in Cupertino, California."
Identify all possible relations (e.g., works-for, headquartered-in) between named entities and represent them as triples. [5]

10. (a) Explain the role of passage retrieval in QA systems. [4]
(b) Describe two traditional methods for passage retrieval and discuss their limitations. [6]
(c) Explain how modern embedding-based retrieval works and its advantages over traditional methods. [5]

11. (a) Define aspect-based sentiment analysis and explain how it differs from traditional sentiment analysis. [3]
(b) Given the review: "The camera quality is amazing, but the battery drains quickly," identify the aspects and their corresponding sentiment polarity. [4]
(c) Explain the typical pipeline of an ABSA system. Include steps like aspect extraction, sentiment detection, and aggregation. [4]
(d) Discuss at least two challenges specific to ABSA and suggest potential solutions. [4]

*** END OF PAPER ***

~~204, 205~~, ~~206~~, 304, 305, 306, 307, 309,