

Reg No.: _____

Name: _____

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY
B.Tech Degree S6 (S,FE) Examination December 2025 (2019 Scheme)

Course Code: AMT302

Course Name: CONCEPTS IN NATURAL LANGUAGE PROCESSING

Max. Marks: 100

Duration: 3 Hours

PART A

Answer all questions, each carries 3 marks.

- | | Mark
s |
|---|-----------|
| 1 Discuss the NLP tasks | (3) |
| 2 Explain the challenges associated with NLP. | (3) |
| 3 Utilize the Bag-of-Words approach to represent the given document: "John likes to watch movies, Mary likes movies too". | (3) |
| 4 Explain the terms Term Frequency and Inverse Document Frequency. | (3) |
| 5 Differentiate Part-of-Speech(POS) tagging and Named Entity Recognition(NER) | (3) |
| 6 In a corpus of 1,000 documents, the term "machine learning" appears in 200 documents. Consider a specific document containing 150 words, where "machine learning" appears 5 times. Calculate the TF-IDF score for the term "machine learning" in this document. | (3) |
| 7 Explain the evaluation process of a Relation Analysis System. | (3) |
| 8 Explain the Vector space model for information retrieval. | (3) |
| 9 Describe the process of rank ordering documents in a collection concerning a query. | (3) |
| 10 List the components constituting a noisy channel model of statistical Machine Translation. | (3) |

PART B

Answer one question from each module, each carries 14 marks.

Module I

- | | |
|--|------|
| 11 a) Explain the Language building blocks of NLP. | (10) |
| b) Explain Heuristic-based NLP. | (4) |

OR

- | | |
|--|------|
| 12 a) Summarize Bayes' Theorem and outline the classification steps with the Naïve Bayes Classifier. | (10) |
| b) Define and discuss the necessity of a kernel function | (4) |

Module II

- 13 a) Explain the stages of the NLP pipeline. (10)
- b) Explain the following terms. (4)
- a) Corpus
 - b) Vocabulary
 - c) Document
 - d) Word

OR

- 14 a) Calculate the TF-IDF representation for the given corpus (10)
- D1:"the house had a tiny little mouse",
 D2:"the cat saw the mouse",
 D3:"the mouse ran away from the house",
 D4:"the cat finally ate the mouse",
 D5:"the end of the mouse story".

- b) Identify the data acquisition methods employed in NLP. (4)

Module III

- 15 a) Given the following training set, train a naive Bayes classifier to perform sentiment analysis on the test sentence: "Predictable with no fun". (10)

Category	Sentence
-	just plain boring
-	entirely predictable and lacks energy
-	no surprises and very few laughs
+	very powerful
+	the most fun film of the summer

- b) Explain Sentiment Analysis with Optimized Naive Bayes Classification. (4)

OR

- 16 a) Explain Name Entity Recognition using Sequence Labelling with an example. (8)
- b) Explain NER, its functioning, and the process of transforming this structured problem into a sequential one akin to POS tagging, where each word has one label. (6)

Module IV

- 17 a) Identify the lightly supervised approaches utilized in NLP. (8)
- b) Describe the supervised approach to relation analysis and outline its limitations. (6)

OR

- 18 a) Given the following data about documents and contents. Create an inverted index for 4 sample documents below. (10)

Doc 1	Sweet sweet nurse! Love
Doc 2	Sweet sorrow
Doc 3	How sweet is love?
Doc 4	Nurse!

- b) Explain Term weighting and document scoring. (4)

Module V

- 19 a) Explain the features to be considered in the passage revival of a factoid question-answering system. (10)
- b) Explain Decoding for Phrase-based Statistical Machine Translation with an example. (4)

OR

- 20 a) Explain a statistical algorithm for word alignment in Machine Translation. (8)
- b) Explain the concept of Mean Reciprocal Rank. (6)
