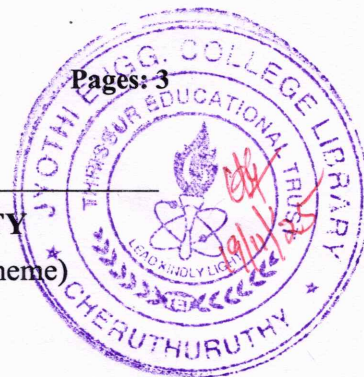


Reg No.: _____

Name: _____

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY
B.Tech Degree S7 (R,S) Examination November 2025 (2019 Scheme)



Course Code: CST473

Course Name: NATURAL LANGUAGE PROCESSING

Max. Marks: 100

Duration: 3 Hours

PART A*Answer all questions, each carries 3 marks.*

Marks

- | | | |
|----|---|-----|
| 1 | Explain morphemes and phonemes with appropriate examples. | (3) |
| 2 | For each sentence, identify whether the different meanings arise from structural ambiguity, semantic ambiguity or pragmatic ambiguity?
1. <i>Time flies like an arrow</i>
2. <i>Visiting relatives can be annoying.</i> | (3) |
| 3 | With appropriate example, explain how stemming is performed. | (3) |
| 4 | Explain sentence segmentation. Mention issues associated with sentence segmentation. | (3) |
| 5 | Explain challenges in Named Entity Recognition. | (3) |
| 6 | Describe any three common applications of text classification in real-world scenarios? | (3) |
| 7 | State the structure of an inverted index used in an information retrieval system. | (3) |
| 8 | Compare information extraction and information retrieval. | (3) |
| 9 | What are factoid questions. Give an example. | (3) |
| 10 | Describe interlingual machine translation approach. | (3) |

PART B*Answer any one full question from each module, each carries 14 marks.***Module I**

- | | | |
|-------|---|-----|
| 11 a) | Explain the building blocks of language with suitable examples. | (9) |
|-------|---|-----|

- b) Discuss the challenges involved in an NLP project. (5)

OR

- 12 a) Explain the different approaches to solve an NLP project (9)
- b) How is classification done by Support Vector Machine on linearly separable data (5)

Module II

- 13 a) Explain any three different vectorization approaches of word embeddings as a model for text representation. (9)
- b) Compare word embeddings model with vectorization approaches. (5)

OR

- 14 a) Supposing that a set of social media posts' dataset is available to do sentiment analysis. What pre-processing steps need to be done in order to use the data for generating a language model? Illustrate. (14)

Module III

- 15 a) Analyze the general pipeline of Information Extraction Process with the help of a diagram. (9)
- b) Explain with suitable example the Laplace smoothing and their use in Naïve Bayes for sentiment classification. (5)

OR

- 16 a) Given the following data about movie review and its classification, classify "predictable with no fun" to one of the classes using Naïve Bayes Classifier. (9)

Document	Category
entirely predictable and lacks energy	Negative
no surprises and very few laughs	Negative
just plain boring	Negative
very powerful	Positive
the most fun film of the summer	Positive

- b) Describe logistic regression for text classification. (5)

Module IV

- 17 a) Explain the Supervised Learning Approach and Lightly Supervised Approach for relation analysis allong with their pros and cons. (10)

- b) Illustrate how would a relational analysis system accurately extract entities and relationships from the sentence: "Marie Curie won the Nobel Prize in Physics in 1903." How is a Relational Analysis System evaluated? (4)

OR

- 18 a) With the help of a diagram explain the architecture of an information retrieval system. (8)

- b) Illustrate how term weighting and document scoring is done for indexing in information retrieval? (6)

Module V

- 19 a) Explain the phases of a factoid question answering system. (9)

- b) Compare the direct and transfer translation techniques of a machine translation system. (5)

OR

- 20 a) State the challenges involved in machine translation. (9)

- b) Explain the concept of Mean Reciprocal Rank. (5)
