

B

1000CST473122201

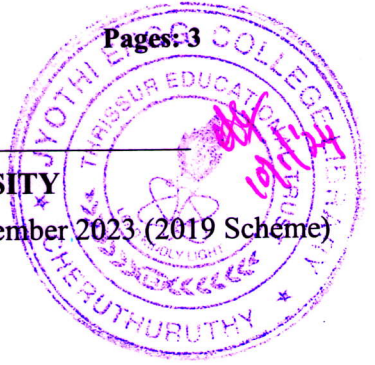
Pages: 3

Reg No.: _____

Name: _____

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

Seventh Semester B.Tech Degree Regular and Supplementary Examination December 2023 (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 | Differentiate information extraction and information retrieval. | (3) |
| 3 | With appropriate example, explain how stemming is performed. | (3) |
| 4 | With the help of a neat diagram explain model stacking and ensembling. | (3) |
| 5 | Identify the named entities for the text:
"Michael Dell is the CEO of Dell Computer Corporation and lives in Austin Texas." | (3) |
| 6 | List any three applications of text classification. | (3) |
| 7 | State the structure of an inverted index used in an information retrieval system. | (3) |
| 8 | Define the terms query, document and collection with respect to an information retrieval system. | (3) |
| 9 | What are factoid questions. Give an example. | (3) |
| 10 | State the challenges involved in machine translation. | (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. | (8) |
| | b) Discuss the challenges involved in an NLP project. | (6) |

OR

- | | | |
|----|---|-----|
| 12 | a) How is classification done by Support Vector Machine on linearly separable data. | (8) |
| | b) Explain the different approaches to solve an NLP project. | (6) |

Module II

- | | | |
|----|--|-----|
| 13 | a) Explain the basic steps of an NLP pipeline. | (8) |
|----|--|-----|

- b) Differentiate Bag of words with Bag of n-grams with the help of an example. Also state the shortcomings of Bag of words model. (6)

OR

- 14 a) Use the toy corpus given below and represent the all the documents using tf-idf model. (8)

Doc 1	Sky gets blue
Doc 2	Sun gets bright.
Doc 3	Sun shines
Doc 4	Sun shines bright

- b) With suitable diagram and example explain the SkipGram model for text representation. (6)

Module III

- 15 a) Given the following data about restaurant review and its classification, classify "Very good food" to one of the classes using Naïve Bayes classification. (10)

Doc 1	Positive	Simply loved
Doc 2	Positive	This place has best food
Doc 3	Positive	Very good restaurant
Doc 4	Negative	Most disgusting food
Doc 5	Negative	Stay away, very disgusting food

- b) Explain with suitable example the use of Laplace smoothing in Naïve Bayes for sentiment classification. (4)

OR

- 16 a) Describe logistic regression for text classification. (8)
 b) Explain the concept of keyphrase extraction. (6)

Module IV

- 17 a) Explain supervised approach for relation analysis. Also state its limitations. (8)
 b) Describe the method used to evaluate the performance of information retrieval system. (6)

OR

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

- b) Explain lightly supervised approach to relation analysis and state its advantages. (6)

Module V

- 19 a) Explain the phases of a factoid question answering system. (10)
b) Compare the direct and transfer translation techniques of a question answering system. (4)

OR

- 20 a) Explain phrase based statistical machine translation system. (10)
b) Explain the concept of Mean Reciprocal Rank. (4)
