

END SEMESTER EXAMINATION - OCTOBER 2025**SEMESTER – 9: INTEGRATED M. Sc. PROGRAMME IN COMPUTER SCIENCE – DATA SCIENCE****COURSE: 21UP9CRMCP30: TEXT ANALYTICS AND NATURAL LANGUAGE PROCESSING***(Regular 2021 Admission))*

Time: Three Hours

Max. Weights: 30

PART A**Answer any 8**

1. State the importance of phrasing ambiguities in Natural Language Processing. (U)
2. Differentiate between list and dictionary in Python with an example for each. (R)
3. State the purpose of Baleen Ingestion Engine. (U)
4. Name the libraries to be imported to perform the following operations:
 - a. Removing stop-words.
 - b. Spell checker
 - c. Lemmatization
 - d. Word Sense Disambiguation
 - e. Regular Expression (A)
5. Define the concept – text classification. (R)
6. Differentiate between NLG and NLU. (U)
7. Define text tokenization, with an example. (Cr)
8. Write a sample script to remove punctuations in a text. (Cr)
9. State any two reasons why models trained on corpus in a specific field or domain of a particular language would perform better than ones trained on general language. Also, list any three examples of publicly available corpus. (An)
10. Given the following pair of sentences, what would be the output after the sentences are preprocessed and a PoS tagging is performed:
Sentence 1: I like to read comic books.
Sentence 2: I read comic book. (A)

(1 x 8 = 8 Weight)**PART B****Answer any 6**

11. Discuss any five challenges faced by NLP applications, quoting an example for each. (An)
12. Explain the different types of stemming with an example script for each. (Cr)
13. Explain the pipeline and the stages involved in text classification. (An)
14. Prepare a brief note on any two types of text similarity checks, along with a sample script for each. (Cr)
15. Write a simple python script that visualizes the frequency count of tokens in a given corpus. (Cr)
16. Explain with examples, on how `regex` is used for text analytics. (Cr)
17. Write the python script that compares the performance of text classifiers with and without cross validation. (Cr)

18. Assuming that a variable contains a short content of your role model, write the python script to perform the following text analytics:

- a. List the set of stop-words as defined in the English language.
- b. Tokenize the given text contained in the variable.
- c. Print the output after removing stop words.

(Cr)

(2 x 6 = 12 Weight)

PART C

Answer any 2 (5 weights each)

19. Natural Language Processing has evolved over time and has created many real-world applications aiding the society in many ways. Explain the significance and evolution of Natural Language Processing.

(An)

20. Explain text classification using Decision Tree utilizing n-grams. Write the python script for the same.

(Cr)

21. Define Context-free grammar. Given the following sentence, decompose the sentence into its grammatical parts:

Sentence: "This boy creates enemies by his attitude."

(An)

22. Assuming that a corpus variable contains a movie description, write python script that decomposes the corpus into its constituent components, and finally tag the tokens according to the Parts-of-Speech.

(Cr)

(5 x 2 = 10 Weight)