(T) Roll No. .....283.....

PAPER ID—11133

B.Tech. EXAMINATION, 2024

(Fifth Semester)

COMPUTER SCIENCE

Specialization Artificial Intelligence

Natural Language Processing

Time : 3 Hours *Maximum Marks : 70*

Before answering the question-paper candidates should ensure that they have been supplied to correct and complete question-paper. No complaint, in this regard, will be entertained after the examination.

Note : Attempt *Five* questions in all, selecting *one* question from each Unit. Q. No. **1** is compulsory. All questions carry equal marks.

1. (a) What is text segmentation, and how does it differ from tokenization ?

(b) What are tagged corpora, and how are they used in NLP ?

(c) How can conditional frequency distributions be used to analyze relationships between word usage and categories ?

(d) What are some common Python control structures used in language processing tasks ?

(e) What are Recurrent Neural Networks (RNNs)

(f) Explain the role of syntax in understanding and processing natural language.

(g) What is a context-free grammar, and how is it used in sentence parsing ? 14

Unit I

2. What is Natural Language Processing (NLP) ? Discuss its goals and challenges in computing with language. How can texts be represented as lists of words in Python ? Explain with an example. What are the advantages of tokenizing text into words for language processing tasks ? 14

3. What is WordNet, and how does it assist in language processing ? Explain, how to access WordNet in Python using the NLTK library ? Provide examples of extracting synonyms and antonyms for a given word. 14

Unit II

4. How can raw text be accessed and processed from web sources using Python ? Illustrate with examples. Discuss methods for reading and writing text files from disk in Python. Provide examples of text processing workflows. **14**

5. What is Unicode, and why is it important in text processing ? Explain, how Python handles Unicode strings ? Provide examples of dealing with non-ASCII characters in text. Discuss the challenges of text processing with multilingual data. How can Python help mitigate these challenges ? **14**

Unit III

6. What is supervised text classification, and how is it applied in natural language processing ? Discuss its key concepts. Explain the workflow of a supervised text classification task. Provide an example using Python for training and testing a classifier. **14**

7. What are the standard metrics for evaluating text classification models ? Discuss precision, recall, F1 score, and accuracy with examples. Explain the importance of confusion matrices in evaluating classification performance. Provide a Python example of generating and interpreting a confusion matrix. **14**

Unit IV

8. What is chunking in NLP, and how does it differ from parsing ? Provide examples. How can a chunker be trained and tested using annotated corpora ? Provide a step-by-step example. What are the key considerations when evaluating the performance of a chunker ? **14**

9. What is relation extraction, and how does it enhance information extraction tasks ? Explain the different types of relations that can be extracted from text, providing examples. What are the challenges in relation extraction, and how can they be addressed using modern NLP techniques ? **14**