6E7108

Total No. of Questions : 22 Total No. of Pages : 04

Roll No. : .....

6E7108

B.Tech. VI-Sem. (Main/Back) Exam. - 2025

ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

6AID5-12 / Natural Language Processing (NLP)

AID, CAI

Time : 3 Hours **Maximum Marks : 70**

Instructions to Candidates :

Attempt all ten questions from Part-A, five questions out of seven questions from Part-B and three questions out of five questions from Part-C.

Schematic diagrams must be shown wherever necessary. Any data you feel missing suitably be assumed and stated clearly. Units of quantities used / calculated must be stated clearly.

Use of following supporting material is permitted during examination.

(Mentioned in Form No. 205)

1. .....NIL..... 2. .....

PART-A **[10x2=20]**

(Answer should be given up to 25 words only)

All questions are compulsory

Q.1. Define Natural Language Processing (NLP).

Q.2. List three common applications of NLP.

Q.3. What is the ArgMax function used for in NLP?

Q.4. Define syntactic collocations.

Q.5. What is term weighting, and why is it important in text processing?

Q.6. What is WordNet?

Q.7. Define Keyphrase Extraction.

Q.8. What is the purpose of PCFGs in NLP?

Q.9. What is Text Clustering?

Q.10. Explain the concept of Sequence Labeling.

PART-B **[5x4=20]**

(Analytical/Problem solving questions)

Attempt any five questions

Q.1. Define the term Machine Learning. Also explain the relationship between Machine Learning and NLP.

Q.2. What is WordNet Similarity? Explain Resnick's work on WordNet similarity measures.

Q.3. What is an IPython Notebook? Discuss the advantages of using IPython Notebooks for NLP projects.

Q.4. What is top-down parsing? Explain the role of parsing algorithms in syntactic analysis.

Q.5. Develop a method of automatically expand user queries semantically related terms from WordNet.

Q.6. What is a Probabilistic Context-Free Grammar (PCFG)? Discuss the training issues associated with PCFGs.

Q.7. Define Graphical Models. Also explain the role of graphical models in sequence labeling.

PART-C [3x10=30]

(Descriptive/Analytical/Problem Solving/Design questions)

Attempt any three questions

Q.1. Describe the role of Part-of-Speech (POS) tagging in syntactic analysis. Analyze the impact of POS tagging accuracy on the performance of downstream NLP tasks.

Q.2. Describe the importance of tokenization in text preprocessing. Implement a tokenizer that can handle punctuation, contractions, and special characters in a given text.

Q.3. Develop a script to generate basic statistics (e.g., word count, sentence count, vocabulary size) for a give text collection.

Q.4. Develop a script to automatically extract keyphrases from a given mystery text and rank them by relevance.

Q.5. Design a system that integrates WordNet with a machine learning model to improve WSD performance.

----- × -----