(T)

Roll No. .....

PAPER ID—11135

B.Tech. EXAMINATION, 2024

(Fifth Semester)

COMPUTER SCIENCE

(Specialization Artificial Intelligence)

Text and Web Intelligence

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. Answer the following questions : $2 \times 7 = 14$

(i) What is the role of the Boolean Retrieval model in Information Retrieval ? Explain with an example.

(ii) Define an Inverted Index and explain its importance in Information Retrieval.

(iii) Describe the preprocessing steps involved in data preprocessing for Information Retrieval.

(iv) What is tf-idf ? How is it calculated and what is its significance in text mining ?

(v) Explain Paradigmatic Relation Discovery with an example.

(vi) Discuss the concept of Sentiment Analysis. How is it different from Opinion Mining ?

(vii) What is the PageRank algorithm and how is it used in web analytics ?

Unit I

2. Define Boolean Retrieval. How does the Term Incidence Matrix help in this retrieval model ? 14

3. Explain the concept of Tokenization in data preprocessing. Why is it important in the context of Information Retrieval ? 14

Unit II

4. What are Skip Pointers in the context of Boolean Retrieval and how do they improve query processing ? 14

5. Discuss the significance of Cosine Similarity in Text Mining and its applications. 14

Unit III

6. What are the differences between PLSA (Probabilistic Latent Semantic Analysis) and LDA (Latent Dirichlet Allocation) ? 14

7. Describe the different types of Recommendation Systems. How do they work ? **14**

Unit IV

8. What is Web Scraping ? Discuss its applications in Web Analytics. **14**

9. Explain the concept of Diffusion in Networks. How is it important in understanding the spread of information on the web ? **14**