(NO. OF PAGES: 2)

(FOR THE CANDIDATES ADMITTED **23UCS6E7**

DURING THE ACADEMIC YEAR 2023 ONLY) **REG.NO. :**

N.G.M.COLLEGE (AUTONOMOUS) : POLLACHI

END-OF-SEMESTER EXAMINATIONS : MAY 2026

BSc COMPUTER SCIENCE(SF) **MAXIMUM MARKS: 75**

SEMESTER: VI **TIME : 3 HOURS**

PART - III

23UCS6E7 – INFORMATION RETRIEVAL

SECTION – A (10 X 1 = 10 MARKS)

ANSWER THE FOLLOWING QUESTIONS. (K1)

1. In an information retrieval system, what is the primary purpose of an inverted index?

(a) To store documents in their original order

(b) To map each document to the list of terms it contains

(c) To map each term to the list of documents in which it appears

(d) To remove stop-words and perform stemming

2. Which retrieval model treats documents as a set of words and uses logical operators (AND, OR, NOT) for exact matching?

(a) Vector Space Model (b) Probabilistic Model (c) Boolean Model (d) Extended Boolean Model

3. Which text categorization algorithm is a probabilistic classifier based on Bayes' theorem and assumes strong independence between features?

(a) K-Nearest Neighbors (KNN) (b) Rocchio (c) Agglomerative clustering (d) Naive Bayes

4. Which technique is a local method improve the recall of an information retrieval system, often involving user input?

(a) Stemming (b) Query expansion (c) Relevance feedback (d) Stopword removal

5. Web mining is the process of discovering useful information from _____.

(a) Structured databases (b) Unstructured or semi-structured web data (c) Image files only (d) Social media interactions only

ANSWER THE FOLLOWING IN ONE (OR) TWO SENTENCES (K2)

6. Define Zipf's law.

7. Define TF-IDF

8. Define Precision and Recall.

9. List the usage of Meta Crawlers.

10. List the applications of Web Mining.

(CONTD …. 2)

Ethical paper

SECTION – B (5 X 5 = 25 MARKS)

ANSWER EITHER (a) OR (b) IN EACH OF THE FOLLOWING QUESTIONS. (K3)

11. a) Apply Porter stemmer rules to process the words: "connecting", "caresses", "ponies" for index term creation in an IR system. **(OR)**
    b) Explain with an example how stop-word removal and stemming are applied during basic tokenizing for building an IR index.

12. a) Illustrate the vector space model with TF-IDF weighting and cosine similarity to rank a given set of documents for a query, showing each calculation step.
    (OR)
    b) Write an example query and document set, demonstrate how extended Boolean matching differs from classical Boolean matching in terms of partial relevance and ranking.

13. a) Illustrate the Rocchio classifier to categorize a given text document into two classes using provided term weights. **(OR)**
    b) Apply k-means clustering to partition a set of 5 text documents into 2 clusters based on TF-IDF vectors.

14. a) Describe the applications of Information Filtering.
    (OR)
    b) Discuss the Hubs and Authorities.

15. a) Analyze the Co-reference resolution techniques with suitable example.
    (OR)
    b) Describe the Web Mining methods and discuss the applications.

SECTION – C (5 X 8 = 40 MARKS)

ANSWER EITHER (a) OR (b) IN EACH OF THE FOLLOWING QUESTIONS. (K4 (Or) K5)

16. a) Examine the impact of the web on traditional IR goals, contrasting historical perspectives with modern challenges. **(OR)**
    b) Inference the role of AI in enhancing IR systems, compare statistical text characteristics like Zipf's law with AI-driven techniques beyond thesauri and morphology.

17. a) Distinguish between Boolean, extended Boolean, vector space, and probabilistic retrieval models.
    (OR)
    b) Explain how different text similarity metrics affect ranked retrieval performance on benchmark text collections.

18. a) Compare the advantages and limitations of Naive Bayes versus KNN in text.
    (OR)
    b) Differentiate LSI and PCA differ in dimension reduction for text clustering.

19. a) Examine the Google PageRank in link analysis for web search ranking.
    (OR)
    b) Describe techniques for duplicate detection in web search engines, and their use in spidering and indexing.

20. a) Differentiate between event extraction and relation extraction with suitable examples.
    (OR)
    b) Explain how web mining uses information from NER and event extraction from web sources.

Ethical paper

****