PRESIDENCY UNIVERSITY
BENGALURU

SCHOOL OF ENGINEERING
MID TERM EXAMINATION - APR 2023

Semester : Semester VI -2020
Course Code : CSE2051
Course Name : Sem VI - CSE2051 - Information Retrieval
Program : ISE

Date : 17-APR-2023
Time : 2PM - 3:30PM
Max Marks : 60
Weightage : 30%

Instructions:
(i) Read all questions carefully and answer accordingly.
(ii) Question paper consists of 3 parts.
(iii) Scientific and non-programmable calculator are permitted.
(iv) Do not write any information on the question paper other than Roll Number.

PART A

ANSWER ALL THE QUESTIONS   **(5 X 2 = 10M)**

1. Find the Score of a term if tf= 45

                                                              (CO1) [Knowledge]

2. Form a sparse matrix from following array :
   ```
   3  2  7
   4  1  0
   3  0  0
   ```

                                                              (CO1) [Knowledge]

3. Define inverted index.

                                                              (CO1) [Knowledge]

4. Explain the Retrieval and Ranking framework from IR architecture.

                                                              (CO1) [Knowledge]

5. Write the advantages of the Vector model with examples.

                                                              (CO2) [Knowledge]

PART B

ANSWER ALL THE QUESTIONS   **(4 X 5 = 20M)**

6. In a collection of documents, for a specific term (T1) the term frequency is 510, and the document frequency for T1 is 327. Find the TF-IDF score of that term if the total number of Documents in the collection is 500

                                                              (CO2) [Comprehension]

7. Find the retrieved result for the following query using Boolean Model. Use the term document incidence matrix for evaluating relevant documents.

Terms	Doc1	Doc2	Doc3
Titan	0	1	0
Persian	1	1	1
Assasin	1	0	1
Sinbad	1	1	0

A) Titan AND Assasin
B) (Sinbad OR Persian) AND Titan
C) ! Persian
D) Titan OR Persian OR Assasin

(CO2) [Comprehension]

8. Love is in the air. In this document, term finds the tf of the query term " Love and Air " with respect to Cosine similarity. Remove all stopwords.

(CO1) [Comprehension]

9. What is the need for the tf-idf weighting mode? Describe all formulas for computing different weights, frequency, and score. Theorem's variable should be well explained

(CO1) [Comprehension]

PART C

ANSWER ALL THE QUESTIONS   **(2 X 15 = 30M)**

10.
Compute the cosine similarities between (a) Doc 1 and Doc 2 and (b) Doc 3 and 4. Term frequency is given for documents.

Terms	Doc1	Doc2	Doc3	Doc4
Information	15	30	20	10
Retrieval	45	6	10	10
Systems	0	15	20	0
Storage	20	25	0	0

(CO2) [Application]

11.
Compute the Cosine similarity between the query "rising star" and the document term "Power of a rising star" when the total documents in the collection are 1200 and Df of the terms are as follows:

Terms	Df
Power	350
Rising	450
Star	535

(CO1) [Application]