

Roll No



**PRESIDENCY UNIVERSITY
BENGALURU**

**SCHOOL OF ENGINEERING
END TERM EXAMINATION - JUN 2023**

Semester : Semester IV - 2021

Course Code : CSE2051

Course Name : Sem IV - CSE2051 - Information Retrieval

Program : ISE

Date : 12-JUN-2023

Time : 9.30AM - 12.30PM

Max Marks : 100

Weightage : 50%

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. How to measure inverse document frequency. (formulae). (CO1,CO2,CO3,CO4)[Knowledge]
- 2. What is User-Based Collaborative Filtering? (CO3,CO4,CO1,CO2)[Knowledge]
- 3. Define the user's task and discuss its relation with the IR problem. (CO2,CO1,CO4,CO3)[Knowledge]
- 4. What are the features of the web crawler? (CO2,CO4,CO3,CO1)[Knowledge]
- 5. What is Recommender Systems. (CO1,CO2,CO3,CO4) [Knowledge]

PART B

ANSWER ALL THE QUESTIONS

(5 X 10 = 50M)

- 6. Compute the Cosine similarity between the query "Gossiping Mammoth" and the document term "Mammoth is found of Gossiping " when the total documents in the collection are 1200 and Df of the terms are as follows:

Terms	Df
Mammoth	350
Gossiping	450
found	535

(CO4,CO3,CO2) [Comprehension]

- 7. Consider the following:
 - i. Neatly draw a web graph by considering the following steps:
 - a. The graph should have 4 nodes (A, B, C, D)
 - b. There is one outdegree from A.
 - c. There are two outdegrees from B.
 - d. There are two indegrees to C.
 - ii. Compute Hubs and authorities of all nodes of the web graph that is constructed using the above steps. (CO2,CO3) [Comprehension]

8. D1: BITS Pilani Goa Campus.
 D2: Presidency University.
 D3: The IIT Delhi.
 -Q: BITS Pilani.

Document frequency

Terms	DF
BITS	30
Pilani	45

In the above 3 documents and one query, find the relevant document by using the cosine similarity. (CO3,CO4,CO2) [Comprehension]

9. Consider the table of term frequencies for 3 documents denoted Doc1, Doc2, and Doc 3 in Fig 1 below. Compute the tf-idf weights for the terms car, auto, insurance, and best for each document using idf values given in Fig 2.

Terms	DOC1	DOC2	DOC3
car	27	4	24
auto	3	33	0
insurance	0	33	29
best	14	0	17

Figure 1

Terms	idf_t	idf_t
Car	18165	1.65
Auto	6723	2.08
Insurance	19241	1.62
Best	25235	1.5

Figure 2

(CO3,CO2) [Comprehension]

10. Consider a document containing 100 words wherein the word Dog appears 3 times, find Term Frequency(TF). Now, assume we have 10 million documents and the word Dog appears in one thousand times of these, find Inverse Document Frequency(IDF) and TF-IDF.

(CO3,CO2) [Comprehension]

PART C

ANSWER ALL THE QUESTIONS

(2 X 20 = 40M)

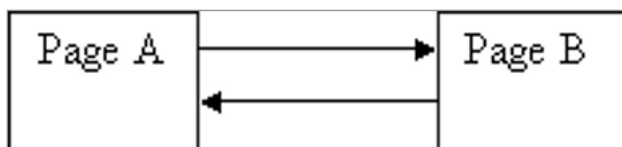
11. A Document collection contains $N = 20$ documents, including:
- d1: "Presidency College"
 - d2: "Presidency University campus"
 - d3: "Presidency Campus"
 - The query is: "Presidency University Campus"
 - User has indicated $R = 6$ relevant documents for this query
 - Query terms: $t_1 = \text{"Presidency"}$, $t_2 = \text{"University"}$, $t_3 = \text{"Campus"}$
 - Document frequencies of query terms in relevant documents and overall collection is given as follows:
 - $rt_1 = 3$, $Nt_1 = 13$
 - $rt_2 = 4$, $Nt_2 = 15$
 - $rt_3 = 2$, $Nt_3 = 12$

To find the relevant document among the three documents, Use the Binary Independence Model(BIM) for the following cases:

- without relevance and without log.
- with relevance and without log

(CO3) [Application]

12. Evaluate the Page Rank of the below graph using the Page Rank Algorithm. Damping Factor(D)=0.70 and $N=2$



(CO4,CO3) [Application]