

Roll No																			
---------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--



**PRESIDENCY UNIVERSITY
BENGALURU
SCHOOL OF ENGINEERING**

MAKEUP EXAMINATION- JAN 2023

Course Code: CSE232
Course Name: Information Retrieval and Organization
Program : B.TECH

Date: 30-JAN-2023
Time: 09:30 AM-12:30 PM
Max Marks: 100
Weightage: 50%

Instructions:

(i) Read the all questions carefully and answer accordingly.

Part A [Memory Recall Questions]

Answer all the Questions. Each question carries FIVE marks. (5Qx 5M= 25M)

1. Difference Between Data and Information Retrieval. (C.O.No.1) [Knowledge's level]
2. Find The Precision for, the given search to the system retrieves 80 items, out of which 40 are relevant and 40 are non-relevant, the precision is 50%. (C.O.No.2) [Knowledge's level]
3. What is KNN? Write the Euclidean Distance Formula for KNN. (C.O.No.3) [Knowledge's level]
4. Define Search Engine. List out the types of Search Engines. (C.O.No.4) [Knowledge's level]
5. What is Collaborative Filtering? List out the types of collaborative filtering. (C.O.No.5) [Knowledge's level]

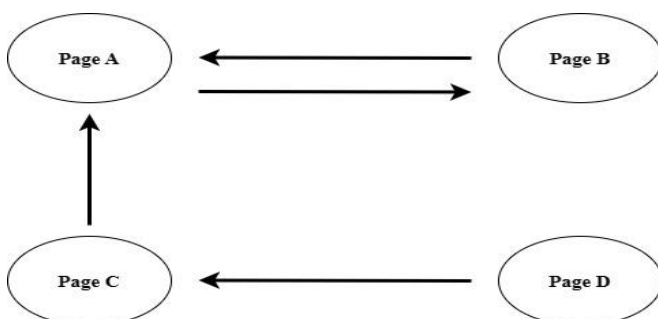
Part B [Thought-Provoking Questions]

Answer all the Questions. Each question carries TEN marks. (3Qx10M=30M)

6. D1: CDs Cheap Software Cheap CDs.
 D2: Cheap Thrills DVDs.
 Q: Cheap CDs Cheap DVDs Extremely Cheap CDs.
 In the above two documents and one query, find the relevant document by using the cosine similarity. (C.O.No.2) [Comprehensive level]
7. Explain in detail about Brute Force Method in the following example. Find the pattern found position.

K	A	R	N	A	T	A	K	A
N	A	T	A	K	A			

8. Find the Page Rank below graph. Damping Factor(D)=0.76 (C.O.No.4) [Comprehensive level]



Part C [Problem Solving Questions]

Answer all the Questions. Each question carries FIFTEEN marks. (3Qx15M=45M)

9. a) Explain the Information Retrieval(IR) with neat architecture describing each component in detail. (10)
b) Write a Short note on Web Crawler and List out the Challenges of the Web. (5)
(C.O.No.1) [Application level]

10. Document collection consists of the following documents

- d1: "Frodo and Sam stabbed orcs"
- d2: "Sam chased the orc with the sword"
- d3: "Sam took the sword"
- The query is: "Sam stabbed orc"
- To find the relevant document among the three documents, Use the Binary Independence Model(BIM) without relevance. (C.O.No.2) [Application level]

11. We have data from the questionnaires survey (to ask people's opinions) and objective testing with two attributes (acid durability and strength) to classify whether a special paper tissue is good or not.

- Here are four training samples.

Name	Acid Durability	Strength	Class
Type 1	7	7	Bad
Type 2	7	4	Bad
Type 3	3	4	Good
Type 4	2	4	Good

Now the factory produces a new paper tissue that passes laboratory tests with $X_1 = 4$ and $X_2 = 7$. Without another expensive survey, can we guess what the classification of this new tissue is using KNN Classifier?

(C.O.No.3) [Application level]