Roll No. ☐☐☐☐☐☐☐☐☐☐☐☐

PRESIDENCY UNIVERSITY

BENGALURU

End - Term Examinations – MAY 2025

Date: 27-05-2025   **Time:** 09:30 am – 12:30 pm

School: SOCSE	Program: ISE	
Course Code : CSE2051	Course Name: Information Retrieval	
Semester: VI	Max Marks:100	Weightage: 50%

CO - Levels	CO1	CO2	CO3	CO4
Marks	26	26	24	24

Instructions:

   (i) Read all questions carefully and answer accordingly.

   (ii) Do not write anything on the question paper other than roll number.

Part A

Answer ALL the Questions. Each question carries 2marks.   **10Q x 2M=20M**

1.	Define the term "Information Retrieval"	2 Marks	L1	CO1
2.	State the main idea behind the Boolean Retrieval Model.	2 Marks	L1	CO2
3.	What is an inverted index?	2 Marks	L1	CO3
4.	Define a recommender system.	2 Marks	L1	CO4
5.	What is the retrieval process in an IR system?	2 Marks	L1	CO1
6.	What is the vector space model?	2 Marks	L1	CO2
7.	Define signature files in text indexing.	2 Marks	L1	CO3
8.	What is a baseline predictor in collaborative filtering?	2 Marks	L1	CO4
9.	Differentiate between data retrieval and information retrieval.	2 Marks	L1	CO1
10.	Define TF-IDF.	2 Marks	L1	CO2

Part B

Answer the Questions.   **Total Marks 80M**

11.	a.	Explain the historical development of IR systems and their modern transformations.	10 Marks	L2	CO1
		Or			
12.	a.	Design a small-scale IR system for an online library and explain its components.	10 Marks	L2	CO1
13.	a.	Documents:	10 Marks	L3	CO1

		D1: "retrieval models are powerful"			
		D2: "retrieval techniques and evaluation"			
		D3: "evaluation of information systems"			
		a) Create a binary term-document matrix b) Represent the query "retrieval evaluation" as a binary vector c) Compute cosine similarity between the query and each document d) Rank the documents based on similarity			
		Or			
14.	**a.**	A collection has the following document word counts: D1: 300 tokens D2: 500 tokens D3: 200 tokens a) Estimate the total token count and vocabulary size (use Heaps' Law) b) If the stopwords are removed, estimate the percentage reduction assuming 30% of vocabulary are stopwords	**10 Marks**	**L3**	**CO1**

15.	**a.**	Compare Boolean, Vector Space, and Probabilistic IR models with real-world use cases.	**10 Marks**	**L2**	**CO2**
		Or			
16.	**a.**	Explain relevance feedback mechanisms with examples of user-based adjustment.	**10 Marks**	**L2**	**CO2**

17.	**a.**	Given Confusion Matrix: TP = 45, FP = 5 FN = 15, TN = 35 Compute: Precision, Recall, F1 Score, Accuracy, Specificity	**10 Marks**	**L3**	**CO2**
		Or			
18.	**a.**	Ranked documents have relevance scores: [3, 2, 3, 0, 1] Ideal ranking = [3, 3, 2, 1, 0] a) Compute DCG b) Compute IDCG for ideal order c) Calculate NDCG	**10 Marks**	**L3**	**CO2**

19.	**a.**	Explain how multilingual crawling is implemented using language detection.	**10 Marks**	**L2**	**CO3**
		Or			
20.	**a.**	Explain crawler politeness and robots.txt protocol.	**10 Marks**	**L2**	**CO3**

21.	a.	Given a mini web graph with 5 pages and the following links:	10 Marks	L3	CO3
		P1 → P2, P3			
		P2 → P4			
		P3 → P2, P4			
		P4 → P5			
		P5 → none			
		a) Perform BFS traversal starting from P1 b) Create the frontier queue at each step c) Simulate indexing after crawling each page			

Or

22.	a.	Given:	10 Marks	L3	CO3
		Query: "search data"			
		D1: "search engine data processing"			
		D2: "retrieval and indexing of data"			
		D3: "data science and analytics"			
		a) Create binary vectors based on term presence b) Compute cosine similarity between query and each document c) Rank documents based on similarity score			

23.	a.	Explain a recommender for an e-learning platform and suggest how feedback is used.	10 Marks	L2	CO4

Or

24.	a.	Explain how explainability improves user trust in recommendations.	10 Marks	L2	CO4

25.	a.	Given item ratings:	10 Marks	L3	CO4

User	Item1	Item2	Item3
U1	4	5	?
U2	3	5	4
U3	2	3	5

a) Compute cosine similarity between Item X and Item Z, Item Y and Item Z

b) Predict U1's rating for Item Z

Or

26.	a.	User U1 has rated:	10 Marks	L3	CO4
		Movie A: Action, Adventure			
		Movie B: Action, Sci-Fi			

		Movie C: Action, Comedy Use binary vectors and cosine similarity to: a) Construct user profile vector b) Compute similarity between profile and Movie C c) Recommend or not?			