Roll No.												

PRESIDENCY UNIVERSITY

BENGALURU

Mid - Term Examinations JUNE-2025

Date: 03- 06-2025 **Time:** 09:30 am – 11:00 pm

School: SOIS	**Program**: MCA	
Course Code : CSA4034	**Course Name:** Information Retrieval and Recommended Systems	
Semester: II	**Max Marks:** 50	**Weightage:** 25%

CO - Levels	CO1	CO2	CO3	CO4	CO5
Marks	14	36			

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. **5Q x 2M=10M**

1	What is the difference between an IR system and a Recommendation system?	**2 Marks**	**L1**	**CO1**
2	What is a dynamic inverted index and why is it needed?	**2 Marks**	**L1**	**CO1**
3	What is the main assumption of the Probabilistic Retrieval Model?	**2 Marks**	**L1**	**CO2**
4	What is a language model in the context of IR?	**2 Marks**	**L1**	**CO2**
5	How does collaborative filtering differ from content-based filtering?	**2 Marks**	**L2**	**CO2**

Part B

Answer the Questions. **Total Marks 40M**

6.	**a.**	Explain the major phases of an index life cycle with an example.	**10 Marks (5+5)**	**L2**	**CO1**
	b.	Explain how index updates are handled in real-time systems.		**L2**	**CO1**
		Or			
7.	**a.**	Describe the steps involved in constructing an inverted index with an example.	**10 Marks (5+5)**	**L2**	**CO1**
	b.	Explain the differences between online and offline index construction?		**L2**	**CO1**

8.	**a.**	How is the relevance of a document estimated in a probabilistic retrieval system?	**15 Marks (5+5)**	**L2**	**CO2**
	b.	Describe the role of relevance feedback in probabilistic models.		**L2**	**CO2**
		Or			
9.	**a.**	Explain the fusion in IR and how does it improve retrieval results?	**15 Marks (5+5)**	**L2**	**CO2**
	b.	Differentiate between early fusion and late fusion techniques.		**L2**	**CO2**

10.		**Solve the following document collection and query:** **D1: "machine learning and artificial intelligence"** **D2: "deep learning for data science"** **Query Q: "learning science"** **The corpus has the following statistics:** **| Term | P(t|C) | |----------|--------| | learning | 0.10 | | science | 0.05 |** **Term Frequencies:** **Term f(t, D1) f(t, D2)** learning 1 1	**15 Marks (5+5+5)**	**L3**	**CO2**

		science 0 1 Document Lengths: D1: 5 terms D2: 5 terms Total terms in collection: 50 **Use Jelinek-Mercer Smoothing with λ = 0.6.** • **Compute the smoothed probability P(Q|D1)** • **Compute the smoothed probability P(Q|D2)** • **Which document is ranked higher based on the query?**			
		Or			

11.		**Given**
		
		• **Document D:** `"information retrieval is essential"`
		• **Query Q:** `"information retrieval system"`
		• **Collection C (Corpus):**
		  • Total number of terms in corpus: 100
		  • Term frequencies in corpus:
		  o information: 10
		  o retrieval: 15
		  o system: 5
		• **Term frequencies in Document D (length = 4):**
		  • information: 1
		  • retrieval: 1
		  • system: 0
		
		Solve the smoothed probability of the query Q using Jelinek-Mercer smoothing with λ = 0.7.
			15 Marks	**L3**	**CO2**