

PRN No.	
---------	--

PAPER CODE	V314-215-C (P)
------------	----------------

(AY:2024-25) December 2024 (ENDSEM) EXAM
TY (SEMESTER - I)

COURSE NAME: INFORMATION STORAGE AND RETRIEVAL **Branch: AI&DS** **COURSE CODE: ADUA31205 C**

(T.Y PATTERN 2020 R1)

Time: [1Hr 30 Min]

[Max. Marks: 40]

(*) Instructions to candidates:

- 1) **Figures to the right indicate full marks. Use required of scientific calculator is allowed**
- 2) **Use suitable data wherever**
- 3) **All questions are compulsory. Solve any two sub question each from Questions 1 and 2**
- 4) **Solve any one sub question (2 marks) from Questions 3 ,4 ,5 and 6 and sub question of 4 marks is compulsory from questions 3,4,5,and 6**

Q. No.	Question Description	Max. Marks	CO mapped	BT Level
Q.1	a) List and describe the main components of an Information storage and retrieval (ISR) system. How do these components interact to provide users with relevant information?	[4]	CO1	Remember
	b) What are the different types of information retrieval models, and what are the major problems encountered in information retrieval?	[4]	CO1	Remember
	c) Discuss the relationship between retrieval and mining. How do these processes complement each other in the context of handling large datasets and extracting useful information?	[4]	CO1	Remember
Q2	a) Define information representation. Why is it important in the context of Information Retrieval systems?	[4]	CO2	Understand
	b) What are pre-processing techniques in Information Retrieval? Explain their role and importance.	[4]	CO2	Understand
	c) Explain the concept of an inverted index in Information Retrieval. How does it work? Consider a corpus with the following collection of 3 documents: 1. D1: "AI in healthcare" 2. D2: "Advances in AI technology" 3. D3: "Healthcare technology advances" Create an inverted index for the given corpus.	[4]	CO2	Understand
Q3	a) How do probabilistic models estimate the relevance of documents to a given query?	[2]	CO3	Understand
	OR			
	b) Identify and explain the primary components of search engine architecture.	[2]	CO3	Understand

	<p>c) The corpus C consists of the following three documents:</p> <p>d1: "new york times" d2: "new york post" d3: "los angeles times"</p> <p>calculate the tf-idf scores for all the terms in C. Assume the words in the vectors are ordered alphabetically. Given the following query: "new new times", calculate the tf-idf vector for the query, and compute the score of each document in C relative to this query, using the cosine similarity measure. Determine which document is the most relevant.</p>	[4]	CO3	Apply
Q4	<p>a) What is precision in the context of information retrieval, and how is it calculated?</p> <p style="text-align: center;">OR</p> <p>b) How does the K-Nearest Neighbors (KNN) algorithm work in text classification?</p>	[2]	CO4	Understand
	<p>c) What is PageRank? How is PageRank calculated? Suppose there is a small web consisting of 4 pages: A, B, C, and D. The link structure between these pages is as follows:</p> <ul style="list-style-type: none"> • Page A links to Page B and C. (On Page A, there are hyperlinks that direct users to Page B and Page C.) • Page B links to Page C. • Page C links to Page A. • Page D links to all pages (A, B, C). <p>Determine the ranking of these pages based on their importance using the PageRank algorithm</p>	[2]	CO4	Apply
Q.5	<p>a) How do relevance models help improve retrieval performance?</p> <p style="text-align: center;">OR</p> <p>b) How can user profiles facilitate personalized search results?</p>	[2]	CO5	Understand
	<p>c) Describe how the Rocchio feedback method can be applied to modify user queries in an Information Retrieval system.</p>	[4]	CO5	Apply
Q.6	<p>a) How does parallel information retrieval improve search efficiency compared to traditional methods?</p> <p style="text-align: center;">OR</p> <p>b) What is web scraping, and how is it typically performed?</p>	[2]	CO6	Understand
	<p>c) Discuss the challenges and techniques in Multimedia Information Retrieval</p>	[4]	CO6	Understand