BIRLA INSTITUTE OF TECHNOLOGY, MESRA, RANCHI
(END SEMESTER EXAMINATION)

CLASS: BTECH **SEMESTER : VI**
BRANCH: AIML **SESSION : SP/2025**

SUBJECT: AI317 INFORMATION RETRIEVAL

TIME: 3 Hours **FULL MARKS: 50**

INSTRUCTIONS:
1. The question paper contains 5 questions each of 10 marks and total 50 marks.
2. Attempt all questions. Mention Section
3. The missing data, if any, may be assumed suitably.
***4*. Before attempting the question paper, be sure that you have got the correct question paper.**
5. Tables/Data hand book/Graph paper etc. to be supplied to the candidates in the examination hall.

			CO	BL
Q.1(a)	Why does a search engine require a logging module? A search engine logs only the top 10% most frequent queries for analytics. It processes 2 million queries per day, with each logged query taking 0.4 KB. a. How many queries are logged daily? b. What is the total storage required per day and for 60 days?	[5]	1	K2
Q.1(b)	Explain the importance of crawling limits and politeness in web crawling. Discuss how web crawlers can be designed to implement these features to avoid overloading a website.	[5]	1	K2
Q.2(a)	Discuss the role of MapReduce in distributed indexing. How does it improve indexing performance?	[5]	2	K2
Q.2(b)	Show how Variable Byte Encoding works step by step with examples, and point out when and why it's preferred over Gamma Encoding.	[5]	2	K2
Q.3(a)	Explain the importance of the Inverse Document Frequency (IDF) in the TF-IDF weighting scheme, and what role does it play in improving the performance of information retrieval systems?	[5]	3	K2
Q.3(b)	Consider a query and a document collection consisting of three documents. Rank the documents using vector space model. Assume tf-idf weighing scheme. Query: "gold silver truck" Document Collection: d1: "Shipment of gold arrived in a truck." d2: "Shipment of gold damaged in a fire." d3: "Delivery of silver arrived in a silver truck."	[5]	3	K5
Q.4(a)	Below is a table showing how two human judges rated the relevance of a set of 12 documents to a particular information need (0 = nonrelevant, 1 = relevant). Let us assume that you've written an IR system that for this query returns the set of documents {4, 5, 6, 7, 8}.	[5]	4	K5

Doc ID	Judge 1	Judge2
1	0	0
2	0	0
3	1	1
4	1	1
5	1	0
6	1	0
7	1	0
8	1	0
9	0	1
10	0	1
11	0	1
12	0	1

PTO

a. Calculate the kappa measure between the two judges.
b. Calculate precision, recall, and F1 of your system if a document is considered relevant only if the two judges agree.
c. Calculate precision, recall, and *F*1 of your system if a document is considered relevant if either judge thinks it is relevant

Q.4(b)	Consider an information need for which there are 4 relevant documents in the collection. Contrast two systems run on this collection. Their top 10 results are judged for relevance as follows (the leftmost item is the top ranked search result): System 1: R N R N N N N N R R System 2: N R N N R R R N N N a. What is the MAP of each system? Which has a higher MAP? Does this result intuitively make sense? Wh b. at does it say about what is important in getting a good MAP score? b. What is the R-precision of each system? (Does it rank the systems the same as MAP?)	[5]	4	K5
Q.5(a)	Explain the difference between local and global query expansion methods. How can poor term selection affect retrieval precision?	[5]	5	K4
Q.5(b)	Omar has implemented a relevance feedback web search system, where he is going to do relevance feedback based only on words in the title text returned for a page (for efficiency). The user is going to rank 3 results. The first user, Jinxing, queries for: banana slug and the top three titles returned are: D1: banana slug Ariolimax columbianus D2: Santa Cruz mountains banana slug D3: Santa Cruz Campus Mascot Jinxing judges the first two documents relevant, and the third nonrelevant. Assume that Omar's search engine uses term frequency but no length normalization nor IDF. Assume that he is using the Rocchio relevance feedback mechanism, with $\alpha = \beta = \gamma = 1$. Show the final revised query that would be run. (Please list the vector elements in alphabetical order.)	[5]	5	K3

::::::01/05/2025::::::M