

**BIRLA INSTITUTE OF TECHNOLOGY, MESRA, RANCHI  
(MID SEMESTER EXAMINATION)**

**CLASS: B. TECH  
BRANCH: CS/IT**

**SEMESTER: VII  
SESSION: MO/2022**

**SUBJECT: IT428 INFORMATION RETRIEVAL**

**TIME: 2 HOURS**

**FULL MARKS: 25**

**INSTRUCTIONS:**

1. The total marks of the questions are 25.
  2. Candidates attempt for all 25 marks.
  3. Before attempting the question paper, be sure that you have got the correct question paper.
  4. The missing data, if any, may be assumed suitably.
  5. Tables/Data hand book/Graph paper etc. to be supplied to the candidates in the examination hall.
- 

		CO	BL
Q1 (a)	While browsing through the morning newspaper, you encounter a new word and look up its meaning in the dictionary. Does this constitute IR? Give your reasons.	[2] 1	K2
Q1 (b)	What are the major blocks of the Indexing Phase of a Search Engine? Which of these are also required in the Query phase?	[3] 1	K1
Q2 (a)	Enumerate at least four different type of web crawlers commonly employed for purpose of IR	[2] 1	K1
Q2 (b)	Suggest a mechanism along with the relevant mathematical model to check for the freshness of a page from the perspective of crawling.	[3] 1	K2
Q3 (a)	Calculate Levenstein's distance between "super" and "surpass" and show the corresponding table.	[2] 2	K3
Q3 (b)	Explain the permuterm method for correcting spellings.	[3] 2	K2
Q4 (a)	An incidence matrix is created for a corpus containing 40,000 unique tokens and 800,000 documents. Each token contains 8 characters on average. If an integer takes 4 bytes on a machine and a character takes 1 byte, what is the size of the incidence matrix in GB, assuming it stores the unique tokens along with the token incidence?	[2] 2	K3
Q4 (b)	Write an algorithm to answer a conjunctive Merge Query from an inverted index in sublinear time complexity.	[3] 2	K2
Q5 (a)	A corpus has 1000000 tokens. Assuming $k=40$ and $b=0.5$ , What is the expected size of the vocabulary? $K$ and $b$ have their usual meanings.	[2] 2	K1
Q5 (b)	Prove that the Logarithmic Merging of indexes is $O(T \log T)$ , where $T$ is the number of postings	[3] 2	K2

: : : : : 26/09/2022 : : : : : M