BIRLA INSTITUTE OF TECHNOLOGY, MESRA, RANCHI
(END SEMESTER EXAMINATION MO/SP20)**

CLASS: BTECH **SEMESTER : VII**
BRANCH: CS/IT **SESSION : MO/2022**

SUBJECT: IT428 INFORMATION RETRIEVAL

TIME: 03 HOURS **FULL MARKS: 50**

INSTRUCTIONS:
1. The question paper contains 5 questions each of 10 marks and total 50 marks.
2. Attempt all questions.
3. The missing data, if any, may be assumed suitably.
***4*. Tables/Data handbook/Graph paper etc., if applicable, will be supplied to the candidates**

Q.1(a) Why is an inverted index called "inverted"? [CO-1, K-1] [2]

Q.1(b) Enumerate some potential drawbacks of using an incidence matrix for the purpose of Information Retrieval.[CO-1, K-2] [3]

Q.1(c) Explain how a crawler ensures it always has a fresh copy of a page? [CO-1, K-3] [5]

Q.2(a) How is stemming different from Lemmatization? [CO-2, K-1] [2]

Q.2(b) Provide the Gamma codes for the numbers 0, 24 and 511. [CO-2, K-2] [3]

Q.2(c) You have been given the task to use **Boolean retrieval** to answer the query **"New Delhi"**. What will be the requirements for an IR system to be able to answer such a query? Provide an algorithm to find documents relevant to the query. [CO-1, K-3] **[5]**

Q.3(a) Given the tables below: [2]

	Doc1	Doc2	Doc3
car	27	4	24
auto	3	33	0
insurance	0	33	29
best	14	0	17

Term	Df	Idf
car	18165	1.65
auto	6723	2.08
insurance	19241	1.62
best	25235	1.5

Compute the tf-idf scores for all the terms for all the documents [CO-3, K-1]

Q.3(b) Discuss one variant of the tf-idf scoring system and compare it with the tf-idf model. [CO-3, K-2] [3]

Q.3(c) Explain the Weighted Zone Scoring system with an algorithm. [CO-3, K-3] [5]

Q.4(a) What is defined as the break-even point in an IR evaluation algorithm? [CO-4, K-1] [2]

Q.4(b) Explain the Mean Average Precision (MAP) evaluation method. [CO-4, K-2] [3]

Q.4(c) Assume the following table is available: [5]

docID	Judge1	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

What is the kappa measure between the two judges? If a document is considered relevant only if the two judges mark it 1, what is the precision and recall of your system if it returns docIDs [4,5,6,7,8] as an answer to a query? [CO-4, K-3]

Q.5(a) Define the term "query expansion". [CO-5, K-1] [2]

Q.5(b) Describe some global methods for query expansion and briefly explain how they work. [CO-5, K-2] [3]

Q.5(c) Explain the Rocchio Algorithm with a suitable figure. [CO-5, K-3] [5]

:::::21/11/2022 M:::::