INDIAN INSTITUTE OF INFORMATION TECHNOLOGY MANIPUR
2nd Assessment Examinations, Aug-2024
B. Tech V Semester- CSE/CSE-AID

Course Name: **Information Retrieval**
Date: **03/09/2024**
Duration: **50 mins**

Course Code: **CS3055**
Time: **8.00 AM to 8.50 AM**
Max. Marks: **25 marks**

Instructions:
-There are two parts of questions: Part A and Part B
- Part A consists of MCQ/ Short answer type which is of 10 marks
- Part B consist of long answer type questions which is of 15 marks
- Marks are shown at the corner in the square bracket

Part A [2x5=10]

1. Formula for weighting ~~term~~ log frequency?
   a) 1 + log(tf)
   b) 1 + log(df)
   c) log(1 + tf)
   d) None
2. What is precision?
   a) TP/TP+FP
   b) TP/TP+FN
   c) FP/TP + FP
   d) None
3. What is recall?
   a) TP/TP+FN
   b) TP/TP+FP
   c) TP/TN+FN
   d) None
4. What is the formula of IDF?
   a) Log (N/DF_t)
   b) Log (N/TF_t)
   c) Log (TF/DF_t)
   d) None
5. Consider a document in a collection containing five different words. A possible language model for that collection might be (0.4, 0.15, 0.25, 0.2, 0.1), where each number is the probability of a word occurring. Let the five words be "Manipur", "north", "state", "east", "part". Calculate the occurrence of the sequence "north east" and "state Manipur".

Part B [5x3=15]

6. Consider the following
   Query = "Newtonian physics"
   The number of documents, N = 10000
   Newtonian occurs in 4500 documents and physics occurs in 200 documents. The document length is 80% of the average length. Assume the tuning parameters k1 = 1.2, b=0.75, k2 = 100 and k = 1.11.
   For a particular document, we are scoring, the query word "Newtonian" occurs 10 times and "physics" occurs 15 times. Calculate the score of the document for the given query using Okapi method.

7. (i) Consider the table of term frequency given below:

Item	Doc1	Doc2	Doc3
Rain	100	5	25
Cloud	70	3	35
Dog	0	60	80

Calculate the cosine similarity of the given documents.

OR

(ii) What is performance measure of an information retrieval system. Explain the various metric to evaluate a retrieval system.

8. (i) Calculate the precision, recall, f-measure and accuracy considering the below table of matrix.

	Relevant	Not Relevant
Retrieved	23 TP	16 FP
Not Retrieved	7 FN	4 TN

OR

(ii) Explain the probability ranking principle and Baye's Rule for document ranking.