

Dr. Arun Kumar Yadav

41/5/23

(E)

(74)

National Institute of Technology Hamirpur
End Semester Examination: 2022-23
Information Retrieval (CS-440)

Branch: B.Tech(CSE)

Time: 2:30 PM – 5:30 PM

Duration: 3 Hrs

Semester: 8th

Date: 03/05/2023

Max. Marks = 50

Note: Attempt all questions from question no. 1 to 5

1. (a) Using bi-grams, which of the following terms is more likely to be considered “equivalent” to the term “informational”? Justify your answer to measure similarity of each term to “informational”.
 - i. informal
 - ii. formalization(b) Discuss the soundex algorithm with suitable example. What are the soundex code for the term “Robert” and “Reupert”. [5+5 = 10 M]
2. Consider the Table 1 below showing how two users rated the relevance of a set of 12 documents to a particular information need (0=non-relevant and 1=relevant). Assume that you have developed an IR system that for this query returns the set of documents (4,5,6,7,8).
 - (a) Calculate the precision, recall and F-measure of your system if a document is considered relevant when users are agreed.
 - (b) Calculate the precision, recall and F-measure of your system if a document is considered relevant when either user it is relevant. [5+5=10 M]

Table 1: Table for Question 2

Doc-id	1	2	3	4	5	6	7	8	9	10	11	12
User-1	0	0	1	1	1	1	1	1	1	0	0	0
User-2	0	0	1	1	0	0	0	0	1	1	1	1

3. What is the use of Latent Semantic Indexing? Explain the evaluation of SVD for LSI with suitable example. [10 M]
4. (a) What are the requirements of text search algorithms? Explain KMP algorithm with suitable example.
(b) What is Probabilistic relevance feedback? Explain probability ranking principle with suitable example. [5+5=10 M]
5. (a) What are the challenges in XML IR? Explain the vector space representation of XML IR with suitable example.
(b) What are the uses of Crawler? Discuss the architecture of crawler with example. [5+5=10M]