R18

Code No: 155EN

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD

B. Tech III Year I Semester Examinations, January/February - 2023

INFORMATION RETRIEVAL SYSTEMS

(Common to CESE, CSE(AIML), CSE(DS))

Time: 3 Hours **Max. Marks: 75**

Note: i) Question paper consists of Part A, Part B.
ii) Part A is compulsory, which carries 25 marks. In Part A, Answer all questions.
iii) In Part B, Answer any one question from each unit. Each question carries 10 marks and may have a, b as sub questions.

PART – A

(25 Marks)

1.a) Define recall. [2]
b) What are the search capabilities of an IDS? [3]
c) What is meant by public index? [2]
d) What is the basis for concept indexing? [3]
e) What is logarithmetic term frequency? [2]
f) List the steps in the clustering process. [3]
g) What is the impact of relevance feedback on search? [2]
h) What is statistical system binding? [3]
i) List the functions supported by Fast Data Finder. [2]
j) What are the five elements of finite state automata used in text searching algorithms? [3]

PART – B

(50 Marks)

2. Describe the item normalization process of information retrieval system in detail. [10]

OR

3.a) Discuss the limitations of term masking.
b) Compare natural language queries with multimedia queries. [5+5]

4.a) Illustrate the two processes associated with information extraction.
b) Demonstrate multimedia indexing. [5+5]

OR

5.a) Make a comparison of dictionary look-up stemmers and successor stemmers.
b) How to create a PAT tree? Explain with example data. [5+5]

6. Explain the need and importance of weighting scheme for automatic indexing and the problems associated with the weighting scheme. [10]

OR

7. Consider the following term-term matrix:

	T1	T2	T3	T4	T5	T6
T1		15	6	8	12	14
T2	15		12	10	6	8
T3	6	12		16	4	10
T4	8	10	16		9	4
T5	12	6	4	9		13
T6	14	8	10	4	13	

a) Determine the Term Relationship matrix using a threshold of 10 or higher
b) Determine the clusters using the clique technique
c) Determine the clusters using the star technique where the term selected for the new seed for the next star is the smallest number term nor already part of a class. [2+4+4]

8.a) Compare and contrast Jaccard measure with Dice measure for similarity.
b) Discuss the significance of negative feedback in ranking the documents. [5+5]

OR

9.a) Explain the potential ambiguities in use of relevance feedback on hypertext documents.
b) Briefly describe the aspects of the visualization process. [5+5]

10. Demonstrate Boyre-Moore Algorithm for the following scenario, explain each step.
String to be searched: abcac
Input String: ababdcabcdacabcac [10]

OR

11.a) Discuss the predominant features of still imagery that can be used in content based indexing.
b) Describe the features of Sagebook for graph retrieval. [5+5]

---ooOoo---