Nirma University

Institute of Technology

Semester End Examination (IR), December - 2021

B. Tech. in Information Technology, Semester-VII

IT702 Information Retrieval Systems

Roll / Exam No. | Supervisor's Initial with Date

Time: 2 Hours | Max Marks :50

Instructions:
1. Attempt all questions.
2. Figure to the right indicate full marks.
3. Draw neat sketches wherever necessary.
4. Assume necessary data wherever required, and indicate clearly.

Q.1 Answer the following: [18]

[A] (CLO 3) For a system with four states (A,B,C,D), the state transition probability matrix is given as mentioned below. [8]

	A	*B*	*C*	*D*
A	0.3	0.4	0.2	0.1
B	0.5	0.1	0.2	0.2
C	0.3	0.3	0.3	0.1
D	0.3	0.2	0.3	0.2

1. If the initial state probabilities are [0 0 1 0] then calculate the probability of achieving state A after three steps.
2. What will be the probability of achieving the sequence AABDC?

[B] (CLO 2) Form a TF-IDF weighted document-term matrix from the following boolean matrix. [10]

Document/ Term	t1	T2	T3	T4	T5
D1	1	0	0	1	0
D2	1	1	1	0	1
D3	1	0	0	0	1
D4	0	1	0	1	1

Q.2 Answer the following: [14]

[A] (CLO 3) Define perceptron and show the classification of XOR operation using it. Discuss the issues which arise during the process, and propose approaches to address the problem. [8]

[B] (CLO 1) What do you think: "Stemming helps in efficient search [6]

results or no?" State with justification.

OR

[B] (CLO 1) A set of documents in a corpus are represented in a two-dimensional manner as follows. [6]

A(4,6) B(2,5) C(7,5) D(5,7) E(6,6)

Compute Euclidean distance and Cosine Similarity between each pair of the objects.

Q.3 Answer the following: [18]

[A] (CLO 1) Apply Condorcet method for meta search on following, and determine the combined rank. [8]

Candidate/ Judge	Judge 1	Judge 2	Judge 3	Judge 4	Judge 5
C1	1	2	1	2	1
C2	2	3	4	1	2
C3	3	4	2	3	-
C4	4	1	3	-	-

[B] (CLO 2) Following matrix represents the graph depicting a small network where each node is a webpage and the edge is a link between two webpages. [6]

$$\begin{bmatrix} & A & B & C & D \\ A & 0 & 1/3 & 1/2 & 1 \\ B & 0 & 0 & 0 & 0 \\ C & 1/2 & 1/3 & 0 & 0 \\ D & 1/2 & 1/3 & 1/2 & 0 \end{bmatrix}$$

1. Draw the graph from the above adjacency matrix.
2. Write the formula for obtaining the rank of the page B.

OR

[B] (CLO 2) The precision for an IR system is 0.4. The F score value is 0.48. Compute the recall value for the system. [6]

[C] (CLO 1) Apply merging algorithm to obtain the common set of documents from the following inverted indices. [4]

T1 - D1 --> D3 --> D4 ---> D5 ---> D8
T2 - D2 --> D4 ---> D6 ---> D9

Show all the steps of the algorithm.