# Nirma University

## Institute of Technology

Semester End Examination (IR), February - 2022

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 five states (A,B,C,D,E), the state transition probability matrix is given as mentioned below. [8]

$$\begin{bmatrix} 0.1 & 0.3 & 0.2 & 0.2 & 0.2 \\ 0.5 & 0.1 & 0.3 & 0 & 0.1 \\ 0 & 0.3 & 0.1 & 0.3 & 0.3 \\ 0.2 & 0.2 & 0.2 & 0.2 & 0.2 \\ 0.4 & 0.3 & 0.2 & 0.1 & 0 \end{bmatrix}$$

1. If the initial state probabilities are [0.5 0 0 0 0.5] then calculate the probability of achieving state A after three steps.
2. What will be the probability of achieving the sequence EBCDA?

[B] (CLO 2) From the following document-term matrix, [10]

1. determine the IDF score of every term in the vocabulary. There are five terms in the vocabulary.
2. Obtain the closest pair of documents based on TF-IDF scoring.

| Document/ Term | T1 | T2 | T3 | T4 | T5 |
|---|---|---|---|---|---|
| D1 | 1 | 0 | 0 | 1 | 0 |
| D2 | 1 | 1 | 2 | 0 | 2 |
| D3 | 1 | 0 | 0 | 0 | 2 |
| D4 | 0 | 2 | 0 | 1 | 1 |

**Q.2 Answer the following:** **[14]**

[A] (CLO 3) Among three popular classification methods: (1) naive Bayesian, (2) Neural Networks, and (3) Support Vector Machines, which one is easy to be adapted to the classification of dynamically changing data streams, and how? Which one is difficult? Discuss. [8]

[B] (CLO 1) Support or refuse the following statement with proper justification. "Stemming increases recall of an IR system." [6]

OR

[B] (CLO 1) Give a scenario of an IR system in which the system has [6]
1. 100% precision but recall is not 100%.
2. 100% recall but precision is not 100%.
3. 100% precision and 100% recall both.

**Q.3 Answer the following:** [18]

[A] (CLO 1) Apply Borda method for meta search on following, and determine the combined rank. Compare your ranking with reciprocal ranking and state your observations. [8]

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

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

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

OR

[B] (CLO 2) How do we address link spamming and how to determine the quality of a webpage? Discuss. [6]

[C] (CLO 1) Demonstrate the concept of mutation and crossover using an appropriate example. [4]