

Nirma University
Institute of Technology
Semester End Examination, May 2021
B.Tech. in Computer Science & Engineering
2CSDE53 Information Retrieval Systems

Roll No.
Time: 90 minutes

Supervisor's Signature with Date :
Max. Marks: 40

- Instructions:
1. Attempt all questions.
 2. Figures to the right indicate full marks.
 3. State and make necessary assumptions wherever necessary.
 4. Draw neat sketches wherever required.

1. Answer the following: **[5]**

- (a) Which kind of queries are used in IR systems? Discuss at least five types of queries with one example of each. **[5]**

2. Answer the following: **[5]**

- (a) Compute Eigen values from following matrix: **[5]**

$$\begin{bmatrix} -5 & 2 \\ -9 & 6 \end{bmatrix}$$

OR

- (a) Describe with any one case study the role of Information Retrieval systems for improving the existing scenario for any real life problem. **[5]**

3. Answer the following: **[5]**

- (a) Consider the following corpus. Assume that the corpus is already preprocessed and all words are equally important (only ignore period(.)). **[5]**

Doc 1: I like to watch Cricket.

Doc 2: Watch the movie tonight.

Doc 3: I lost my watch yesterday.

Doc 4: I will watch a movie tomorrow.

Use probabilistic bigram language model using this corpus to calculate the probability of following sentence: *I watch Cricket.*

OR

- (a) For following documents retrieved in response to a query, assume that there are 7 relevant documents as per the ground truth for this query. **[5]**

Rank position	1	2	3	4	5	6	7	8	9
Relevant?	YES	YES	NO	YES	NO	NO	YES	NO	NO

1. (3 marks) Draw P-R curve by listing all values.
2. (2 marks) Determine average precision.

4. Answer the following: **[7]**

- (a) In a singing competition, five contestants are to be judged by three expert jury members. Each contestant is to be given an integer score from 1 to 10. Table 1 shows the scores given to each participant by respective jury member. **[7]**

Use Condorcet method to determine the final combined ranking and choose the winner.

Table 1: Scores given to participants by jury members

Contestant	Jury 1	Jury 2	Jury 3
C1	8	7	5
C2	2	5	7
C3	6	6	6
C4	9	8	8
C5	5	5	9

5. Answer the following:

[8]

- (a) Apply multinomial naive Bayesian classifier on following training data representing documents using bag of words model. Use Laplacian smoothing to avoid zero probability error and predict the class label for the document : “*tense angry sad happy smile*”.

[8]

Document	Bag of words	Label
1	angry sad disgust calm	negative
2	smile smile happy happy	positive
3	angry smile angry smile	neutral
4	sad sorry sorry smile	negative
5	tense calm happy	neutral
6	calm happy smile disgust	positive

6. Answer the following:

[10]

- (a) Consider following four documents:

[10]

D1: Time and tide wait for none.

D2: Time is the best medicine.

D3: Waiting time for medicines is long.

D4: Did you take medicine now?

- (5 marks) Represent this corpus using tf-idf weighting scheme after necessary preprocessing.
- (5 marks) For the query *Time for medicine now*, determine the ranking of the documents using cosine similarity.