

[This question paper contains 4 printed pages.]

Your Roll No.....

Sr. No. of Question Paper : 7291

L

Unique Paper Code : 2343010028

Name of the Paper : Information and Image Retrieval

Name of the Course : **B.Sc. (H) CS**

Semester : VIII

Duration : 3 Hours

Maximum Marks : 90

Instructions for Candidates

1. Write your Roll No. on the top immediately on receipt of this question paper.
2. **Section A** is compulsory (**Question 1**).
3. Attempt **any 4** questions from **Section B**.
4. Parts of a question should be attempted together.

Section A

- Q1 (a) Differentiate between tokenization and stop word removal with suitable examples. 2
- (b) Discuss why the Boolean retrieval model fails to rank results. 2
- (c) Differentiate between color histogram and color correlogram with respect to pixel relationships. 2
- (d) How do texture features differ from color features in representing an image? 3
- (e) Discuss invariant color features and their importance in image retrieval. 3
- (f) Explain the effect of the curse of dimensionality on image retrieval systems. 3
- (g) A document collection contains the words: 3
- connect, connected, connecting, connection, and connections.
- Apply stemming and count the number of unique terms before and after stemming.

P.T.O.

- (h) Given two feature vectors: 4

Query image: $Q = (4, 6)$

Database image: $D = (2, 3)$

The inverse covariance matrix is:

$$S^{-1} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$$

- i. Find the difference vector $Q - D$.
 - ii. Apply the Mahalanobis distance formula.
- (i) Define the concept of the semantic gap in the context of Content-Based Image Retrieval (CBIR). How does Relevance Feedback bridge the semantic gap? 4
- (j) Explain Fourier Descriptors and their use in shape representation. 4

Section B

- Q2 (a) Consider the following documents: 5
- D1: "information retrieval system retrieval"
 D2: "retrieval models and system"
 D3: "information system design"
- i. Compute the term frequency (TF) of the term "retrieval" in D1.
 - ii. Compute the inverse document frequency (IDF) of the term "retrieval".
 - iii. Calculate the TF-IDF value of "retrieval" in D1.
- (b) A system retrieves 20 documents, out of which 12 are relevant. The total number of relevant documents in the collection is 30. 5
- i. Calculate Precision
 - ii. Calculate Recall
 - iii. Compute F-measure ($\beta = 1$)
- (c) Given the following documents: 5

D1: "hotel booking online payment"

D2: "online food delivery payment"

D3: "hotel reservation and booking"

After preprocessing, answer the following:

- i. Write the vocabulary of the document collection.
- ii. Construct the postings lists for the terms hotel and booking.
- iii. Evaluate the Boolean query: hotel AND booking.

- Q3 (a) Given a 4×4 grayscale image with intensity levels 0-2 as follows: 7

$$\begin{bmatrix} 1 & 0 & 1 & 2 \\ 2 & 1 & 0 & 1 \\ 1 & 2 & 1 & 0 \\ 0 & 1 & 2 & 1 \end{bmatrix}$$

Construct the Gray Level Co-occurrence Matrix (GLCM) for $\theta = 0^\circ$ and distance $d = 1$, normalize the GLCM, and compute the Energy (ASM) and Contrast measures.

- (b) Define the Precision-Recall curve. Given the following Precision-Recall values: 8
 (Recall, Precision): (0.2, 0.9), (0.4, 0.8), (0.6, 0.7), (0.8, 0.6). Draw the Precision-Recall curve and discuss the effect on the curve when recall increases but precision decreases.

- Q4 (a) For a query, the total number of relevant images is 3. The system retrieves relevant images at ranks 1, 4, and 8, with $K=8$. Compute the Modified Retrieval Rank (MRR) and the Average Modified Retrieval Rank (AVMRR). 6

- (b) Given two feature vectors $A = [1, 2, 3]$ and $B = [2, 4, 6]$, compute the Euclidean distance, Earth Mover's Distance and the Canberra distance between them. 9

- Q5 (a) Describe the Tamura texture features coarseness, contrast, and directionality. 6

- (b) Compare the CBIR system QBIC, VisualSEEK and WebSEEK in terms of query type, features used, and application suitability. 9

- Q6 (a) Given the feature similarities: Color = 0.9, Texture = 0.6, and Shape = 0.7, with weights 0.5, 0.3, and 0.2 respectively: 5

- i. Compute the weighted similarity score and identify which feature contributes the most.
- ii. Discuss what happens if the texture weight increases.

- (b) Why is dimensionality reduction required in image retrieval systems? What is the main idea of FLDA? Give one limitation of FLDA. 5

- (c) Consider the following document collection: 5

Doc 1: new home sales top forecasts

Doc 2: home sales rise in july

Doc 3: increase in home sales in july

Doc 4: july new home sales rise

- i. Construct the term-document incidence matrix for the above documents.
- ii. Build the corresponding inverted index.
- iii. Which representation is more efficient for searching and why?

- Q7 (a) Compute the LBP value for the center pixel (value = 30) using the following 3×3 neighborhood: 7

55	58	52
49	50	60
48	47	53

- i. Generate the binary pattern.
- ii. Convert it to decimal LBP value.

- (b) An image has 10 pixels with the following color distribution: 8

Red: 6 pixels → 4 are connected, 2 are isolated

Blue: 4 pixels → 3 are connected, 1 is isolated

Threshold for coherence = 3 pixels

Identify coherent and incoherent pixels and Compute the CCV representation