

# INDIAN STATISTICAL INSTITUTE

THIRD SEMESTER EXAMINATION, 2014/16 SESSION

## PAPER - 13

### INFORMATION STORAGE, RETRIEVAL AND DBMS

(MASTER IN LIBRARY INFORMATION SCIENCE)

18<sup>th</sup> November 2015

10:00-13:00

(3 Hours)

---

#### INSTRUCTION TO STUDENT

1. This Question paper consists of 4 pages including coverpage.
2. **Attempt Questions and/or Sub-Questions sufficient enough to score 100 marks.**
3. Please print all your answers in the answer Booklet provided.
4. Scientific Calculator is allowed.

### QUESTION 1.

Write briefly about

- i) Advanced information storage devices [4 marks]
- ii) Role of information retrieval in Big Data Analytics. [4 marks]
- iii) Boolean text retrieval technique [4 marks]
- iv) Role of RDBMS and E-R Diagrams in text information retrieval [4 marks]
- v) Information Retrieval in Library Information Science [4 marks]

### QUESTION 2.

- A. Give a schematic diagram of the system for summarization of multiple documents. [4 marks]
- B. Define (i) Bag of Words, (ii) Inverse Document Frequency (IDF), (iii) Collection frequency, (iv) Document frequency, (v) Absolute frequency, (vi) Recall measure, (vii) a simple algorithm (steps) for string-matching to remove redundant sentences. [12 marks]
- C. Using K- Mixture model, write the formula—with notations involved—used in the calculation of  $P_i(k)$ , probability of word ( $w_i$ ) that appears  $k$  times in a document. [8 marks]

### QUESTION 3.

- A. Figure 1 is a four-bit greyscale image,  $f(x,y)$ , of size 7 x 7 pixels, depicting greyscale values of a picture in a document. These greyscale values range between 0 and 15 (4-bit image).

1	2	7	6	7	8	4
13	11	9	15	3	4	5
12	11	9	14	7	6	5
4	6	14	7	2	6	8
0	1	3	14	11	13	15
1	4	7	8	12	14	16
8	10	14	11	13	11	9

Figure 1

Compute  $\sum_{x,y} f(x,y)$ , and plot a histogram for the above 7 X 7 multi-valued image [6 marks]

- B. What are the three types of image models? Explain Hierarchical Minimal Complexity Measure Threshold (HMCMT) algorithm for properly converting a grayscale image into a binary image. [6 marks]

### QUESTION 4.

- A. What is spatial autocorrelation? Compute the Moran's Index for a spatial field of your choice. Provide all the steps involved in estimating Moran's I. [8 marks]
- B. What is Mahalanobis Distance? Compute it for the data, created of your choice, containing two variables (x and y) belonging to two groups (X and Y). Write your conclusions about the relevance of Mahalanobis Distance in the context of information retrieval. [8 marks]

### QUESTION 5.

- A. Write briefly about (i) Tonal similarity, and (ii) Time and pitch [4 marks]
- B. Let a database consist of a large collection of songs. Such songs can be represented as patterns on two-dimensional space, when the two spatial coordinates—time ( $x$ ), and pitch values ( $y$ )—are available. Explain Hit-Or-Miss Transformation (HMT), and Blur Hit-Or-Miss Transformation (BHMT) and their use in Music Information Retrieval (MIR). [8 marks]

### QUESTION 6.

Why delimiter space of a document is important from the point of its geometric composition? Explain how the geometry of delimiter space could be quantitatively characterized via rectangular granulometries. Give full set of equations involved in such a quantitative characterization. [8 marks]

### QUESTION 7.

Let  $X$ ,  $Y$ ,  $Z$  be three disjointed spatial zones of a spatial system  $S$ . Let  $mX$ ,  $mY$ , and  $mZ$  be variables of corresponding zones, and these variables are of different strengths. In Modified Gravity Model, the required parameters are asymmetric distances, location significance indices. Write full set of equations to classify these three zones in terms of force of attraction between all possible combinations of the zones. Provide interaction matrices in support of the equations. [8 marks]

### QUESTION 8.

Write a full set of equations involved in:

- (i) converting point-data into polygonal data via Skeletonization by Influence Zone (SKIZ), and Weighted Skeletonization by Influence Zone (WSKIZ) transformations [5 marks]
- (ii) Computing morphological median between the sets  $X$  and  $Y$ , where the intersection of  $X$  and  $Y$  yields a nonempty set. [5 marks]
- (iii) Computing grayscale morphological median between the two spatial functions  $f$  and  $g$  such that the point-wise minimum between  $f$  and  $g$  is neither  $f$  nor  $g$ . [5 marks]

### QUESTION 9.

Following are the three spatial fields  $f^1$ ,  $f^2$ , and  $f^3$  of each of size  $5 \times 5$ .

208	25	41	37	168	71	128	192	245	215	44	140	47	237	79
231	72	248	108	10	174	245	66	140	65	166	76	94	198	130
33	140	245	234	217	168	87	130	36	208	187	190	160	125	131
233	245	124	203	239	42	150	179	39	63	166	49	199	112	209
162	247	205	245	174	31	58	228	66	237	115	176	21	114	203

(a)

(b)

(c)

(Contd...)

The dilation  $d(f^i, f^j)$  and erosion  $e(f^i, f^j)$  distances interaction matrices for the above three spatial fields are given below:

$d(f^i, f^j)$				$e(f^i, f^j)$			
	$f^1$	$f^2$	$f^3$		$f^1$	$f^2$	$f^3$
$f^1$	1	5	5	$f^1$	1	4	3
$f^2$	5	1	2	$f^2$	4	1	3
$f^3$	5	2	1	$f^3$	3	3	1

Compute the ranks for all possible pair of spatial fields.

**[10 marks]**

**END OF PAPER**