

Reg. No. :

--	--	--	--	--	--	--	--	--	--

**Question Paper Code: 52932**

M.E. DEGREE EXAMINATION, DECEMBER 2015

Elective

Computer Science and Engineering

15PCS510 - INFORMATION RETRIEVAL TECHNIQUES

(Regulation 2015)

Duration: Three hours

Maximum: 100 Marks

Answer ALL Questions

PART A - (5 x 1 = 5 Marks)

1. Cosine similarity vector is used to find
  - (a) Similarity between 2 term
  - (b) Similarity between 2 documents
  - (c) Similarity between 2 files
  - (d) None of the above
2. Cosine similarity vector is used to find
  - (a) similarity between 2 term
  - (b) similarity between 2 documents
  - (c) similarity between 2 files
  - (d) none of the above
3. What is CPC
  - (a) Cost per Click
  - (b) Cost per Code
  - (c) Cost per Circuit
  - (d) None of the above
4. Bow tie structure of web graph is
  - (a) In and out node
  - (b) in node only
  - (c) out node only
  - (d) none of the above
5. F measure is
  - (a)  $2PR/P+R$
  - (b)  $3PR/P+R$
  - (c)  $2.5 PR/P+R$
  - (d) none of the above

PART - B (5 x 3 = 15 Marks)

6. Distinguish between IR Versus Web Search.
7. Explain the characterization of IR Models.
8. Define permuterm index.
9. List the merits of Boolean retrieval.
10. Define lammatization.

PART - C (5 x 16 = 80 Marks)

11. (a) Apply Leveinstein edit distance between GOOD and GOD. (12)

Or

- (b) (i) Compute variable byte code for Docids 44444 and 33333 applying VB Encoding and decoding algorithms. (8)

- (ii) Compute Huffman code for the following” SHE SELLS THE SEA SHELLS”. (8)

12. (a) Find Kappa Statistics( $K$ ), Precision( $P$ ), Recall( $R$ ) and MAP for the following: (16)

Judge 1	Y	Y	N	Y	Y	N	N	Y	Y
Judge 2	N	Y	N	Y	N	N	N	Y	Y

Or

- (b) Write XML for the following scenario: Book- purchase- from Book shop which contains 3 books , authors , Price, publisher information , buyer information and the total cost of books purchased. (16)

13. (a) Discuss the role of Mercator with neat sketch. And justify the need of URL frontier, DNS and robots exclusion protocol. (16)

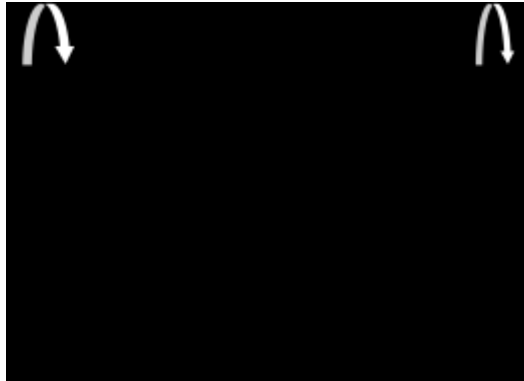
Or

- (b) Consider a web graph with three nodes 1, 2, 3 and 4. The links are as follows: 1 to 1, 1 to 2, 2 to 1, 2 to 3, 3 to 2, 4 to 3, 4 to 2, 3 to 4 and 3 to 1. Write down the transition probability matrices for the surfers walk with teleporting value = 0.5 and find the page rank vector. (16)

14. (a) Analyze the following text and find the fingerprint using simhash technique: “lotus is a flower” (assume the weight values). (16)

Or

(b) Find HITS score for the following (16)



15. (a) Compute  $k$  means clustering for the following documents which contains their docid and score values. Group them in to 2 clusters:  $A_1(5, 5)$ ,  $A_2(8, 8)$ ,  $A_3(9, 9)$ ,  $A_4(10, 10)$ ,  $A_5(11, 11)$  Assume initial seeds are  $A_2$  and  $A_4$ . (16)

Or

(b) Find the near duplicate for the table using Hamming distance problem. ( $F$  = Finger print certificate).  $F = 0100\ 1110$ ,  $k = 3$  bit (16)

Finger Print
1100 0101
1111 1111
1100 0101
1110 0111
1011 1111
0100 1101
1011 0100
0000 1110
1111 0110
0100 1110

