APR-17/BE/Insem-73

B.E. (Information Technology) (Semester -II)
INFORMATION STORAGE AND RETRIEVAL
(2012 - Pattern)

Time : 1 Hour]   *[Max. Marks :30*

Instructions to the candidates:

   1)   *Answer Q1 or Q2, Q3 or Q4, Q5 or Q6*

   2)   *Neat diagram must be drawn wherever necessary.*

   3)   *Figures to the right indicate full marks.*

   4)   *Assume suitable data if necessary*

Q1) a)   Explain working of conflation algorithm in detail. Justify use of this algorithm in information retrieval. **[6]**

    b)   Write a short note on matching coefficients. **[4]**

OR

Q2) a)   Why single pass algorithm is better than Rocchio's Algorithm? **[10]** Form the document cluster of following document term matrix using single pass clustering algorithm.

    Consider

    Membership function : Sum of product

    centroid calculation function : Average

    Threshold = 11

	D1	D2	D3	D4	D5
T1	1	1	0	1	1
T2	2	1	2	3	0
T3	3	0	1	0	1
T4	2	2	0	3	0
T5	2	2	1	2	1

Q3) a) Compare boolean model and vector model. Explain how vector model can be used to retrieve partial matching documents. **[6]**

b) What are inverted files? Explain how these file can be used to answer Boolean queries. **[4]**

<div align="center">OR</div>

Q4) a) Explain working of suffix tree. Construct suffix tree for following example "This is a text. A text has many words. Words are made from letters."**[6]**

b) Explain working of signature files with example? **[4]**

Q5) a) Wirte a note on user oriented measures used for evaluating the performance of any retrieval system. Also explain their significance. **[6]**

b) Explain the term ontology creation **[4]**

<div align="center">OR</div>

Q6) a) Write a note on " Ontology languages for semantic web". **[5]**

b) Explain the trade-off between precision and recall. **[5]**

<div align="center">ⓘ ⓘ ⓘ</div>