FACULTY OF ENGINEERING

B.E. 4/4 (CSE) II – Semester (Main) Examination, May / June 2012

Subject : **Information Retrieval Systems** (Elective – III)

Time : 3 hours

Max. Marks : 75

Note: Answer all questions from Part–A and answer any **FIVE** questions from Part–B.

PART – A (25 Marks)

1. What is the difference between the concept of a "Digital Library" and an information retrieval system?

2. Why should researches in information retrieval care about standards?

3. Does high specificity always imply high exhaustivity. Justify your answer.

4. Apply the porter stemming steps to the word : triplicate.

5. Under what conditions would the Bayesian and the vector approach be the same.

6. What is a dendogram?

7. Is the use of positive feed back always better than using negative feedback to improve a query?

8. What are the two broad classes of visualization?

9. Define Unique-relevance-recall.

10. Why is relevance feedback not used by most search engines?

PART – B (50 Marks)

11.a) What are the four major functional processes of information storage and retrieval system? Describe each of these processes in brief.

 b) What is relationship between vocabulary browse and thesauri/concept classes?

12.a) How does the process of information extraction differ from the process of document indexing?

 b) Describe the similarities and differences between term stemming algorithms and n-grams. Describe how they affect precision and recall.

13.a) Describe how use of natural language processing will assist in the disambiguation process.

 b) Briefly outline the process of manual clustering.

14.a) What are the three different levels of query binding? Why these levels are required in the creation of a search?

 b) Explain the use of tree maps in information visualization.

15.a) Trade off the use of hardware versus software text search algorithms citing advantages and disadvantages of each in comparison to the other.

b) What are the main problems associated with generalizing the results from controlled tests on information systems to their applicability to operational systems.

16.a) Conceptually, what role does Hyper text play in indexing and the definition of an item? What is the major problem with URLs as a basis for hypertext links?

b) Discuss the differences between the single link, the clique and the star clusters.

17. Write short notes on :
   a) PAT data structure
   b) Boyer-Moore algorithm
   c) Concept indexing

- - - - - - -