

B.E.(with Credits)-Regular-Semester 2012-Information Technology Sem VII  
**IT 704-Elective-I : Information Retrieval Systems**

P. Pages : 2

Time : Three Hours



**GUG/W/16/6607**

Max. Marks : 80

- Notes :
1. All questions carry marks as indicated.
  2. Due credit will be given to neatness and adequate dimensions.
  3. Assume suitable data wherever necessary.
  4. Illustrate your answers wherever necessary with the help of neat sketches.

1. a) Explain the total information systems with the help of a neat diagram. **10**
- b) Explain in detail the concept of Precision & Recall. **6**

**OR**

2. a) Why is the concept of processing tokens introduced and how does it relate to a word? What is the impact of searching being based on processing tokens versus the original words in an item? **8**
- b) Describe the rationale why use of proximity will improve precision versus use of just the Boolean functions. Discuss its effect on improvement of recall. **8**
3. a) Explain in detail PORTER's stemming Algorithm. **10**
- b) What is Automatic Indexing & Manual Indexing? Explain the differences. Also mention the advantages & disadvantages of both techniques. **6**

**OR**

4. a) How does the process of information extraction differ from process of document indexing? **6**
- b) Does high specificity always imply high exhaustively? Justify your answer. **5**
- c) Trade off the use of Precordination versus Post coordination. **5**
5. a) Explain Probabilistic Weighting Technique in Automatic Indexing process. **8**
- b) Write a short note on Natural language processing & concept-Indexing for Automatic Indexing. **8**

**OR**

6. a) How do the concepts underlying Discrimination value indexing provide a good or poor basis for searching items? What is the major problem with weight formula? **8**
- b) Discuss how the decision on use of different techniques can affect the user's ability to find information. **8**

7. a) What is clustering? Explain the steps & Guidelines for the process of clustering. 8
- b) Explain manual clustering Techniques with the help of KWOC, KWIC & KWAC. 8

**OR**

8. a) Can statistical Thesaurus Generation be used to develop a hierarchical cluster representation of a set of items? Discuss the process how you would use it in a system. 8
- b) Prove that a term could not be found in multiple clusters when using the single link technique. 8
9. a) Discuss the sources of potential errors in the final set of search terms from when a user first identifies a need for information to the creation of final query. 8
- b) What are some potential ambiguities in use of relevance feedback on hypertext documents? 8

**OR**

10. a) What are search statements? Explain the levels of binding? 8
- b) Discuss the techniques of Hidden Markov Models & write their problems. 8

\*\*\*\*\*