USN | | | | | | | | | | |

10SSE322

Third Semester M.Tech. Degree Examination, December 2012
Information Retrieval

Time: 3 hrs. Max. Marks:100

Note: ***Answer any FIVE full questions.***

1 a. What is meant by information retrieval? Bring out the difference between information retrievals and data retrieval. **(08 Marks)**

b. Give the formal characterization of IR model. Explain the vector model of IR and list its advantages and disadvantages. **(12 Marks)**

2 a. Compare Boolean, vector and probabilistic models, with respect to information retrieval. **(06 Marks)**

b. What is retrieval performance evaluation? Which are the basic measures of retrieval evaluation? Explain single value summaries for retrieval evaluation. **(08 Marks)**

c. Explain query formulations based on pattern matching. **(06 Marks)**

3 a. Discuss the various hierarchical model of structure queries. **(10 Marks)**

b. What is the advantages of user relevance feedback as a query reformulation strategy? Explain how it is used for query expansion with the vector model. **(10 Marks)**

4 a. Discuss the role of clustering in automatic local analysis. Explain the different strategies for building local clusters. **(12 Marks)**

b. Discuss the various issues in modeling natural language. **(08 Marks)**

5 a. Explain query expansion based on a statistical thesaurus. **(10 Marks)**

b. What is the motivation for text compression in information retrieval? Explain the basic approaches to text compression. **(10 Marks)**

6 a. What is an inverted file? Explain the construction of an inverted file with a suitable example and list the steps of the search algorithm for an inverted index. **(12 Marks)**

b. Explain shift – or and suffix automation techniques for sequential searching, with suitable examples. **(08 Marks)**

7 a. What is the motivation for parallel information retrieval? How is partitioned parallel processing accomplished on a MIMD machine? **(06 Marks)**

b. Explain logical document partitioning and physical document partitioning for system that use inverted files. **(08 Marks)**

c. Explain the various issues in the design of a distributed IR system. **(06 Marks)**

8 a. Discuss the various approaches to the use of recommender agents. **(08 Marks)**

b. What are the challenges passed by the web in IR? **(08 Marks)**

c. Briefly explain the PageRank algorithm. **(04 Marks)**

*****