Eighth Semester B.E. Degree Examination, May/June 2010
Information Retrieval

Time: 3 hrs.   Max. Marks:100

Note: *Answer any FIVE full questions, selecting*
at least TWO questions from each part.

PART – A

1   a. What is information retrieval? What are the components of information retrieval systems? Define precision and recall. **(10 Marks)**
    b. Explain how vectors are used to represent documents. With an example show how to construct document vectors and calculate similarity coefficients between the given query and the documents in vector space model. **(10 Marks)**

2   a. Briefly explain inference networks. Describe how inference networks are used for relevance ranking, with an example. **(12 Marks)**
    b. Give an overview of information retrieval based on genetic algorithms. **(08 Marks)**

3   a. What is clustering of documents? What are the steps of hierarchical agglomerative clustering of documents? Explain various hierarchical clustering algorithms. **(10 Marks)**
    b. Write a note on passage based retrieval. **(04 Marks)**
    c. Explain the notion of semantic networks? Explain how 'R-distance' is used for measuring distance between two sets of nodes. **(06 Marks)**

4   a. What are invalid files? Explain with an example. **(04 Marks)**
    b. What are the general steps followed in the search algorithm on an inverted index? **(06 Marks)**
    c. Explain KMP and BM algorithms for text searching. Compare their time complexity. **(10 Marks)**

PART – B

5   a. Explain query translation and document translation strategies for cross language retrieval. **(06 Marks)**
    b. Explain how stemming and transliteration utilities are useful for cross language information retrieval. **(10 Marks)**
    c. Write the pseudocode for I-Match duplicate document detection algorithm. **(04 Marks)**

6   a. Describe the various approaches used to include IR functionality in a relational DBMS. **(12 Marks)**
    b. Explain various methods used for storing XML data and accessing them efficiently. **(08 Marks)**

7   a. Explain the various approaches used to parallelise the scanning of signature files. **(10 Marks)**
    b. Give the brief idea of calculating page ranking on a web search. **(04 Marks)**
    c. Discuss P2P information systems. What are the ranking criteria, used by P2P information retrieval system? **(06 Marks)**

8   a. Discuss in detail, the main aspects considered in designing a multimedia query language. **(10 Marks)**
    b. Give the basic idea behind the GEMINI approach for indexing multimedia objects and write the algorithm for fast similarity searching. **(10 Marks)**

* * * * *