

(2½ Hours)

[Total Marks: 75]

- N.B. 1) All questions are compulsory.  
2) Figures to the right indicate marks.  
3) Illustrations, in-depth answers and diagrams will be appreciated.  
4) Mixing of sub-questions is not allowed.

**Q. 1 Attempt ANY FOUR from the following: (20M)**

- (a) Explain the components of IR System.  
(b) Step by step explain the inverted index construction for the below given documents.  
Document 1: The IR practical was conducted in CS lab.  
Document 2: The CS lab was occupied whole day for a practical.  
(c) Write an algorithm to find Edit Distance. Use Edit Distance algorithm to find the edit distance between “sunday” and “saturday”.  
(d) Write a short note on Vector Space Model.  
(e) Discuss various evaluation metrics used in information retrieval.  
(f) List the different spelling correction techniques. Explain any one in detail.

**Q. 2 Attempt ANY FOUR from the following: (20M)**

- (a) Write a short note on Naive Bayes classification algorithm.  
(b) Diagrammatically explain web search architecture.  
(c) Briefly explain page rank algorithm. Give suitable example to explain the concept.  
(d) Discuss RankBoost.  
(e) State the use of crawler. Explain in brief the working of crawler/ crawler architecture.  
(f) Explain hierarchical clustering.

**Q. 3 Attempt ANY FOUR from the following: (20M)**

- (a) Which various dynamic web contents need to be taken care of during crawling?  
(b) Briefly explain summarization. List the types of summarization and explain any one in detail.  
(c) State the different approaches to find precise answers in Question Answering.  
(d) Discuss Item based collaborative filtering. Provide suitable example.  
(e) Write the challenges and techniques of multilingual retrieval.  
(f) Elaborate the concept of A/B testing.

**Q. 4 Attempt ANY FIVE from the following: (15M)**

- (a) What are the challenges of IR?  
(b) With the help of an example, elaborate the concept of HITS algorithm.  
(c) Briefly explain Surveys in information retrieval.  
(d) Define Boolean model along with a suitable example.  
(e) Write the applications of text categorization.  
(f) State the different web crawling techniques and explain any one in brief.

\*\*\*\*\*