(2 ½ 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)    Define Information Retrieval (IR) and explain its goals.
(b)    How are documents represented in an IR system? Discuss different term weighting schemes.
(c)    Describe the Boolean model in Information Retrieval. Discuss Boolean operators and query processing.
(d)    Write a short note on Vector space model.
(e)    What is edit distance, and how is it used in measuring string similarity? Provide examples.
(f)    What is Experimental design and significance testing.

Q. 2    Attempt **ANY FOUR** from the following: (20M)
(a)    Write a short note on support vector machine.
(b)    Explain K-means clustering technique.
(c)    State the challenges in web search.
(d)    What is the role of crawler in information retrieval? How does it accomplish its task?
(e)    List link analysis algorithms and write the applications of link analysis.
(f)    Discuss pairwise learning.

Q. 3    Attempt **ANY FOUR** from the following: (20M)
(a)    What is text summarization? Explain extractive summarization.
(b)    Discuss web crawling techniques.
(c)    Write a short note on user based collaborative filtering.
(d)    State the techniques and applications of cross-lingual retrieval.
(e)    Briefly explain Surveys in information retrieval.
(f)    Explain Near-duplicate page detection algorithms.

Q. 4    Attempt **ANY FIVE** from the following: (15M)
(a)    Discuss storage mechanisms for indexed documents.
(b)    What are the Techniques used for spelling correction in IR systems.
(c)    Explain Cosine Similarity in Information Retrieval.
(d)    Briefly explain RankBoost Algorithm.
(e)    How to handle dynamic web content during crawling in IR system.
(f)    Write a short note on Test collections and benchmarking.

*****************