(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.

1. Attempt *any four* of the following: **20**

a. What is information retrieval example? What are the characteristics of information retrieval.
b. What are the components and What are the major challenges faced in Information Retrieval.
c. What is edit distance, and how is it used in measuring string similarity with suitable example.
d. Explain the process of constructing an inverted index. How does it facilitate efficient information retrieval?
e. What is relevance feedback in the context of retrieval models.
f. Explain Vector space model. Discuss TF-IDF, cosine similarity.

2. Attempt *any four* of the following : **20**

a. Define text categorization and explain its importance in information retrieval systems.
b. How can clustering be utilized for query expansion and result grouping in information retrieval systems.
c. Explain the effectiveness of K-means and hierarchical clustering in text data analysis.
d. Explain the architecture of a web search engine. What are the components involved in crawling and indexing web pages.
e. What is the role of supervised learning techniques in learning to rank and their impact on search engine result quality.
f. Discuss the difference between the PageRank and HITS algorithms.

3. Attempt *any four* of the following : **20**

a. Explain breadth-first and depth-first Web page crawling Techniques?
b. Define near-duplicate page detection and its significance in web search. Explain the challenges associated with identifying near-duplicate pages.
c. Describe common techniques used in extractive text summarization.
d. What are Challenges associated with question answering.
e. Define collaborative filtering and content-based filtering in recommender systems.
f. Explain different approaches to machine translation, including rule-based, statistical, and neural machine translation models.

4. **Attempt *any five* of the following :** **15**

a Discuss the steps involved in the Soundex Algorithm for phonetic matching.

b Construct 2-gram, 3-gram and 4-gram index for the following terms:
a. banana
b. pineapple
c. computer

c Discuss the Naive Bayes algorithm for text classification. How does it work, and what are its assumptions.

d Discuss how link analysis can be used in social network analysis and recommendation systems.

e Discuss challenges in abstractive text summarization.

f Describe the role of test collections and benchmarking datasets in evaluating IR systems.
