Time : Three hours   Maximum : 100 marks

Answer ALL questions.

PART A — (10 × 2 = 20 marks)

1. Outline the impact of the web on information retrieval.

2. Compare information retrieval and web search.

3. Define an inverted index.

4. What is stemming? Give example.

5. Define web crawling.

6. What is schema heterogeneity?

7. List the characteristics of MapReduce.

8. What do you mean by item-based collaborative filtering?

9. Differentiate between information filtering and information retrieval.

10. What are the desirable properties of a clustering algorithm?

PART B — (5 × 16 = 80 marks)

11. (a) (i) Appraise the history of information retrieval.   (8)

    (ii) Explain the role of artificial intelligence in information retrieval. (8)

Or

    (b) What is search engine? Explain with diagrammatic illustration the
    components of a search engine.   (16)

12. (a) (i) Explain vector space retrieval model with an example. (10)

      (ii) Write short notes on query expansion. (6)

Or

(b) (i) What is relevance feed back? Explain with an example an algorithm for relevance feedback. (8)

      (ii) Give an example for latent semantic indexing and explain the same. (8)

13. (a) (i) Explain the working of a web crawler with an example. (10)

      (ii) Explain index compression with an example. (6)

Or

(b) Explain with an example the framework of a XML retrieval system. (16)

14. (a) (i) Write short notes on topic specific page rank computation. (4)

      (ii) Explain the HITS link analysis algorithm with an example. (12)

Or

(b) Explain collaborative filtering and content based recommendation system with an example. (16)

15. (a) (i) State Bayes theorem. (4)

      (ii) Explain Naïve bayes classification with an example. (12)

Or

(b) What is clustering? Explain k-means clustering algorithm with an example. (16)

———————