Reg. No. : | | | | | | | | | | | | |

Question Paper Code : 80426

B.E./B.Tech. DEGREE EXAMINATIONS, APRIL/MAY 2024.

Sixth/Seventh/Eighth Semester

Computer Science and Engineering

CS 8080 – INFORMATION RETRIEVAL TECHNIQUES

(Common to : Computer and Communication Engineering/ Information Technology)

(Regulations 2017)

Time : Three hours Maximum : 100 marks

Answer ALL questions.

PART A — (10 × 2 = 20 marks)

1. Define Precision and Recall with respect to Information retrieval.

2. List the challenges in designing efficient information retrieval system.

3. How will you calculate the TF and IDF values of a word in a text document?

4. Give the use of retrieval metrics.

5. What do you understand by the term, "Curse of Dimensionality".

6. State the advantages of using SEO.

7. Define hub score and authority score.

8. List few applications of a web crawler.

9. Define: Recommender system.

10. Can content-based filtering be used in all applications? Justify.

PART B — (5 × 13 = 65 marks)

11. (a) With a neat sketch, elaborate about the software architecture of an information retrieval system.

Or

(b) (i) Explore how the proliferation of web search engines has influenced information seeking behavior and the practice of information retrieval. (6)

(ii) Explain about the components of a search engine. (7)

12. (a) Consider a document collection containing 500 documents and a vocabulary of 100 unique terms. Elaborate on the steps in building a vector space model (VSM) and a probabilistic model for this collection.

Or

(b) Consider a document collection containing 500 documents and a vocabulary of 100 unique terms. Assume that you are given a task of improving retrieval performance of document collection. Elaborate on the steps with relevance feedback and query expansion techniques.

13. (a) Ram maintains a dataset consisting of 1000 emails, with 600 labeled as spam and 400 labeled as not spam. He wants to create probabilistic model for predicting whether the new incoming email is spam or not. Elaborate the steps to implement the same.

Or

(b) You are tasked with comparing the performance of sequential searching and multidimensional indexing techniques for searching a dataset of 1,00,000 records. Each record contains four attributes: ID, Name, Age, and Salary. Explain about the application of sequential searching and multidimensional indexing over this dataset with 1,00,000 records.

14. (a) List the uses of HITS algorithm. Elaborate about its implementation procedure.

Or

(b) Explain the Collaborative Filtering with clustering technique in detail.

15. (a) With a neat diagram, explain the process of content-based filtering mechanism.

Or

(b) With a neat diagram, explain the process of matrix factorization-based recommendation system.

PART C — (1 × 15 = 15 marks)

16. (a) Web crawlers are used to systematically traverse the World Wide Web and collect web pages for indexing. This system needs to be scalable, efficient, and capable of handling millions of web pages to support the millions of users.

(i) With a neat diagram, elaborate about the components of web crawler. (10)

(ii) Discuss the various challenges incurred while implementing an efficient web crawler. (5)

Or

(b) Describe in detail about the working of IR architecture with a neat sketch. Elaborate about the ranking process.