

binils.com & binils android app
Photoplex.net & android app

78

Reg. No. :

Question Paper Code : 20399

B.E./B.Tech. DEGREE EXAMINATIONS, APRIL/MAY 2022.

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 Information Retrieval.
2. Compare and contrast Information Retrieval and Database Retrieval.
3. Give the taxonomy of Information Retrieval models.
4. What do you mean by Relevance Feedback?
5. Mathematically define the Text Categorization Problem.
6. Discuss about Curse of Dimensionality.
7. Outline the differences between Information Retrieval and Web Search.
8. Define Recall and Precision in the context of a Search Engine.
9. What is a Recommender System? Explain.
10. Give an utility matrix for an example of your choice

PART B — (5 × 13 = 65 marks)

11. (a) Give the architecture of a typical IR system describing each component in detail. (13)

Or

- (b) Discuss in detail the challenges of Web Search and a detailed note on search interfaces. (13)

Question Paper Sponsored by
CSI Institute OF Technology, Thovalai, Kanyakumari Dist.

12. (a) Given the following documents, create the vector space model, compute the Inverse document frequency, compute the weight of the words and normalize all documents to unit length.

Doc1 : Text Mining is an important area of research. Research in Text Mining involves machine learning.

Doc2 : Text classification and text clustering are important areas of text mining. Other areas include feature extraction and dimensionality reduction.

Doc3 : There are many algorithms for text classification. Supervised techniques of machine learning are generally used for text classification.

Doc4 : There are many algorithms for text clustering. Unsupervised techniques of machine learning are generally used for text clustering. (13)

Or

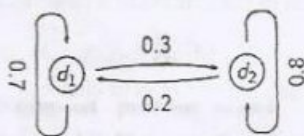
- (b) Compare and contrast Boolean models, Vector Space models and Probabilistic Models of IR. (13)

13. (a) Explain in detail about Naïve Bayes algorithm and its application in Text classification. (13)

Or

- (b) Discuss in detail about SVM Classifier and their use in Text Classification. (13)

14. (a) Draw the architecture of the Google search engine and elaborate. Compute the PageRank for the given graph. (13)



Or

- (b) What is Link analysis? How is it used for Ranking Pages? Explain the HITS algorithm. How is it different from PageRank algorithm? (13)

PhotoPlex. Net & android app

binils.com & binils android app

binils.com & binils android app
Photoplex.net & android app

80

15. (a) Explain Collaborative Filtering. Discuss the neighborhood based algorithm used for collaborative filtering. (13)

Or

- (b) Given the following Matrix :

	Lion King	Aladdin	Mulan	Anastasia
John	3	0	3	3
Joe	5	4	0	2
Jill	1	2	4	2
Jane	3	?	1	0
Jorge	2	2	0	1

Find Jane's rating of Alladin using user-based collaborative filtering explaining the steps in detail. (13)

PART C — (1 × 15 = 15 marks)

16. (a) You are asked to design an internal search engine for your organization. Explain in detail the steps you would take and the various factors you would consider for designing without any flaws. (15)

Or

- (b) You are asked to design a text classification engine to process all queries raised by the employees of your organization. Elaborate in detail about the steps you will take and the various factors to be considered while designing.