

Reg. No. :

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Question Paper Code : 50410

B.E./B.Tech. DEGREE EXAMINATIONS, APRIL/MAY 2023

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. What are the impacts of information retrieval on the web?
2. List out the issues in information retrieval systems.
3. Construct a vector space model representation.
4. Interpret cosine similarity measure.
5. What is hash based dictionary in indexing?
6. Differentiate range queries and nearest neighbor queries.
7. What are the challenges in data traversing by search engines?
8. Classify the use of Web indexing and inversion of indexing process.
9. Point out some advantages of Mobile Recommendation system.
10. Construct Collaborative and Content based recommendation system.

PART B — (5 × 13 = 65 marks)

11. (a) (i) Compare in detail about Information Retrieval and Web Search with examples. (8)
(ii) Summarize the functions and features of Information Retrieval Systems. (5)
- Or
- (b) (i) Demonstrate the working of IR architecture with a diagram. (6)
(ii) Infer how designing parsing and scoring functions works in detail(7)

12. (a) (i) Discuss in detail about term frequency and Inverse Document Frequency. (7)

(ii) Compute TF-IDF given a document containing terms with the given frequencies A(3), B(2), C(1). Assume document collections 10,000 and document frequencies of these terms are A(50), B(1300), C(250). (6)

Or

(b) (i) Explain about latent semantic indexing and latent space with an illustration. (7)

(ii) Illustrate the probabilistic relevance feedback, pseudo relevance feedback and Indirect relevance feedback. (6)

13. (a) (i) Evaluate agglomerative clustering and HAC in detail. (7)

(ii) Summarize the significance of SVM classifier in detail (6)

Or

(b) Examine single dimensional and multi-dimensional indexing with an example.

14. (a) Design and develop a Web search Architecture and the components of search engine and its issues.

Or

(b) Describe the following with example.

(i) Bag of Words and Shingling (7)

(ii) Hashing, Min Hash and Sim Hash (6)

15. (a) (i) Explain about Matrix Factorization? (4)

(ii) Discuss the approaches of recommender system. (9)

Or

(b) (i) Write down the common HLA terminologies (4)

(ii) Define the steps involved in Collaborative Filtering (9)

PART C — (1 × 15 = 15 marks)

16. (a) The following lists of Rs and Ns represents relevant (R) and nonrelevant (N) returned documents in a ranked list of twenty documents retrieved in response to a query from a collection of 10,000 documents. The top of the ranked list (the document the system thinks is most likely to be relevant) is on the left of the list. This list shows six relevant documents. Assume that there are eight relevant documents in total in the collection.

RRNNN NNNRN RNNNR NNNNR

- (i) What is the precision of the system on the top twenty? (3)
(ii) What is the F1 on the top twenty? (3)
(iii) What is the uninterpolated precision of the system at 25% recall? (3)
(iv) Assume that these twenty documents are the complete result set of the system. what is the MAP for the query? (3)
(v) Assume now that the system returned the entire 10,000 documents in a ranked list and these are the first twenty results returned. What is the largest possible MAP that this system could have? (3)

Or

- (b) (i) Rank the impacts of Categorization and clustering of text in the mining with the suitable examples. (8)
(ii) Formulate the application of Near Duplicate Document Detection techniques and also Generalize the advantages in Plagiarism checking. (7)