

Reg. No.																			
----------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Question Paper Code	13934
---------------------	-------

B.E. / B.Tech. - DEGREE EXAMINATIONS, NOV / DEC 2025

Seventh Semester

Computer Science and Engineering

(Common to Information Technology)

20CSEL703 - INFORMATION RETRIEVAL TECHNIQUES

Regulations - 2020

Duration: 3 Hours

Max. Marks: 100

PART - A (MCQ) (10 × 1 = 10 Marks)

Answer ALL Questions

	<i>Marks</i>	<i>K- Level</i>	<i>CO</i>
1. The fundamental problem addressed by Information Retrieval systems is: (a) Storing large amounts of data efficiently (b) Interpreting the meaning of user queries (c) Finding material of an unstructured nature that satisfies an information need (d) Converting data into information through processing	1	K1	CO1
2. A key difference between Information Retrieval and Data Retrieval is that Information Retrieval primarily deals with: (a) Structured records in a database (b) Unstructured or semi-structured data (c) Providing a single, exact answer to a query (d) High precision at the cost of recall	1	K1	CO1
3. In the context of the Vector Space Model, what does the TF-IDF weighting scheme aim to achieve? (a) Calculate the probability of a document's relevance (b) Identify the importance of a term in a document relative to a collection (c) Model the latent semantic relationships between terms and documents (d) Execute logical operations for document filtering	1	K1	CO2
4. A retrieval system returns 8 documents for a query. Out of these, 5 are relevant. The total number of known relevant documents in the entire collection for that query is 10. What is the Recall of the system? (a) 0.50 (b) 0.625 (c) 0.50 (5/10) (d) 1.25	1	K1	CO2
5. The primary purpose of an 'Inverted Index' in text searching is to: (a) Reduce the dimensionality of the feature space (b) Map terms to the documents that contain them for efficient retrieval (c) Visualize the clustering of similar documents (d) Calculate the accuracy of a classification model	1	K1	CO3
6. In a text classification task, if a model correctly classifies 87 out of 100 documents, what is its Accuracy and Error rate, respectively? (a) Accuracy: 87%, Error: 13% (b) Accuracy: 0.87, Error: 0.13 (c) Accuracy: 13%, Error: 87% (d) Accuracy: 0.13, Error: 0.87	1	K1	CO3
7. The primary purpose of a Scheduling Algorithm in a web crawler is to: (a) Rank the crawled pages by their relevance (b) Determine the link-based importance of a page (c) Decide the order in which to download pages from the crawl frontier (d) Distribute the crawling load across a cluster of machines	1	K1	CO4
8. Learning to Rank techniques in search engines differ from simple ranking functions because they: (a) Rely exclusively on the analysis of the link structure between web pages (b) Are used for scheduling the order of web crawling (c) Use machine learning to combine multiple ranking signals into a single relevance score (d) Are only applicable in distributed search engine architectures	1	K1	CO4

9. The core function of a Content-Based Recommender System is to suggest items that are: 1 K1 CO5
 (a) Popular among all users in the system
 (b) Similar to the items a user has liked in the past
 (c) Recommended by users with a similar taste
 (d) Newly added to the system's inventory
10. The primary goal of applying principles of cognition and perception to information visualization in a web search engine is to: 1 K1 CO6
 (a) Increase the raw processing speed of the search algorithm
 (b) Present search results in a way that allows users to understand and use the information more effectively
 (c) Store larger amounts of indexed data
 (d) Automatically reformulate the user's query for better precision

PART - B (12 × 2 = 24 Marks)

Answer ALL Questions

11. List any four key components of the Software Architecture of an IR System. 2 K1 CO1
12. Differentiate between 'Data Retrieval' and 'Information Retrieval' based on the nature of the data and the output. 2 K2 CO1
13. Define the two core components of the TF-IDF weighting scheme. 2 K1 CO2
14. Compare Precision and Recall as retrieval metrics, stating what aspect of performance each one measures. 2 K2 CO2
15. State the fundamental difference between Supervised and Unsupervised algorithms in the context of text learning. 2 K1 CO3
16. Explain why Feature Selection or Dimensionality Reduction is a critical step before applying a classification algorithm like SVM on text data. 2 K2 CO3
17. List the four main components typically found in the architecture of a web crawler. 2 K1 CO4
18. Extend the core idea behind Link-based Ranking and provide the name of a well-known algorithm that uses this principle. 2 K2 CO4
19. List any two common Data and Knowledge Sources that can be used as input for building a Recommender System. 2 K1 CO5
20. Interpret the difference of the core principles of Content-Based Filtering and Collaborative Filtering. 2 K2 CO5
21. Find the role of cognition in the context of designing visualizations for a web search engine. 2 K1 CO6
22. How does an understanding of preattentive processing in human perception influence the design of a search engine results page? 2 K1 CO6

PART - C (6 × 11 = 66 Marks)

Answer ALL Questions

23. a) Explain the 'Retrieval and Ranking Processes' in an IR System. 11 K2 CO1
OR
 b) Demonstrate the components of Information Retrieval and Search engine. 11 K2 CO1
24. a) Develop the Vector space retrieval model with example. 11 K3 CO2
OR
 b) Construct a Latent semantic indexing and latent semantic space with an example. 11 K3 CO2
25. a) Compare and contrast the k-Nearest Neighbors classifier and the Support Vector Machine classifier for text classification and Outline the basic principle of each algorithm and discuss a scenario where one would be preferred over the other. 11 K2 CO3
OR
 b) Show the process of creating and using an Inverted Index for efficient text searching. 11 K2 CO3

26. a) Explain the Cluster-based and Distributed architectures for a search engine. 11 K2 CO4
OR
b) Illustrate the working of scheduling algorithms in Web Crawler. 11 K2 CO4
27. a) Extend the high-level architecture of a Content-Based Recommender System. 11 K2 CO5
OR
b) Compare and contrast Neighborhood models and Matrix Factorization models used in Collaborative Filtering. 11 K2 CO5
28. a) Model the visualization for query specification. 11 K3 CO6
OR
b) Identify the techniques used for interactive visualization. 11 K3 CO6