Reg. No.																

Question Paper Code	13215

B.E. / B.Tech. - DEGREE EXAMINATIONS, NOV / DEC 2024

Seventh Semester

Computer Science and Engineering

20CSEL703 - INFORMATION RETRIEVAL TECHNIQUES

Regulations - 2020

Duration: 3 Hours | Max. Marks: 100

PART - A (MCQ) (20 × 1 = 20 Marks)

Answer ALL Questions

		Marks	K – Level	CO
1.	Which of the following is not a component of an information retrieval system? (a) Query processing (b) Indexing (c) Ranking (d) Data analytics	1	K1	CO1
2.	A group of related documents against which information retrieval is employed is called as (a) Corpus (b)Text Database (c)Index Collection (d) Repository	1	K1	CO1
3.	How the information retrieval problem can be defined formally? (a) a triple (b) a quadruple (c) a couple (d) None of the above	1	K1	CO1
4.	The similarity between two vectors in the Vector Model is usually calculated using (a) Jaccard Index (b) Cosine similarity (c) Euclidean distance (d) Manhattan distance	1	K1	CO2
5.	The count of occurrences of a word in a document is referred as (a) document frequency (b) term frequency (c) collection frequency (d) change frequency	1	K1	CO2
6.	In Boolean retrieval, each item in the list which records that the term is appeared in the document is called as (a)Ranking (b) Posting (c) Indexing (d) Grepping	1	K1	CO2
7.	Which of the following best describes text classification? (a) Grouping similar texts (b) Assigning predefined labels to text data (c) Extracting keywords from text (d) Generating summaries of text	1	K1	CO3
8.	The Decision tree algorithm works best with (a) Numerical data only (b) Categorical and numerical data (c) Categorical data only (d) Time series data	1	K1	CO3
9.	Which algorithm is known for being a "lazy learner"? (a) Decision Tree (b) KNN (c) SVM (d) Naive Bayes	1	K1	CO3
10.	How does divisive hierarchical clustering differ from agglomerative clustering? (a) Divisive starts with one cluster and splits, while agglomerative starts with individual points and merges (b) Divisive only works for categorical data (c) Agglomerative requires fewer calculations than divisive clustering (d) Divisive is faster than agglomerative clustering	1	K1	CO4
11.	In agglomerative clustering, how are clusters merged? (a) From the largest to the smallest (b) By progressively merging the nearest clusters (c) By dividing the largest cluster into smaller clusters (d) Randomly merging clusters until K is reached	1	K1	CO4
12.	Which method of unsupervised learning involves grouping similar data points together? (a) Clustering (b) Regression (c) Classification (d) Prediction	1	K1	CO4
13.	Which architecture allows multiple servers to work together to handle search requests? (a) Centralized Architecture (b) Cluster-based Architecture (c) Distributed Architecture (d) Peer-to-Peer Architecture	1	K1	CO5
14.	Which ranking algorithm is particularly effective in handling large datasets? (a) TF-IDF (b) PageRank (c) BM25 (d) Cosine similarity	1	K1	CO5

15.	Which of the following best describes the process of web crawling? (a) Fetching pages from the web (b) Analyzing user behavior (c) Storing data in a database (d) Generating user queries	*1*	*K1*	*CO5*
16.	Which metric is commonly used to evaluate search engine performance? (a) Click-through rate (b) User engagement time (c) Recall and precision (d) Server response time	*1*	*K1*	*CO5*
17.	What is the primary function of a recommender system? (a) Categorize products (b) Predict user preferences (c) Generate advertisements (d) Store data	*1*	*K1*	*CO6*
18.	What type of recommender system uses user profiles to make predictions? (a) Collaborative filtering (b) Content-based filtering (c) Hybrid filtering (d) Matrix factorization	*1*	*K1*	*CO6*
19.	What is the primary drawback of content-based filtering? (a) Cold start problem for new users (b) It does not consider item features (c) High computational cost (d) Limited to already known preferences	*1*	*K1*	*CO6*
20.	In a hybrid recommendation system, the purpose is to (a) Combine multiple techniques to improve accuracy (b) Only use content-based filtering (c) Only use collaborative filtering (d) Avoid neighborhood models	*1*	*K1*	*CO6*

PART - B (10 × 2 = 20 Marks)

Answer ALL Questions

21.	Write note on logical view of the document.	*2*	*K1*	*CO1*
22.	Compare data retrieval and information retrieval.	*2*	*K2*	*CO1*
23.	What is Probabilistic Information Retrieval?	*2*	*K1*	*CO2*
24.	What is Latent Semantic Indexing?	*2*	*K1*	*CO2*
25.	Define Supervised learning.	*2*	*K1*	*CO3*
26.	What are support vectors in a SVM model?	*2*	*K1*	*CO3*
27.	What is PCA?	*2*	*K1*	*CO4*
28.	What is meant by feature selection?	*2*	*K1*	*CO4*
29.	Define the URL frontier in a web crawler.	*2*	*K1*	*CO5*
30.	What is the role of user ratings in collaborative filtering?	*2*	*K1*	*CO6*

PART - C (6 × 10 = 60 Marks)

Answer ALL Questions

31.	a)	Describe the various components of Information Retrieval System with neat a diagram.	*10*	*K2*	*CO1*
		OR			
	b)	Discuss in detail on document Pre-Processing in Information Retrieval.	*10*	*K2*	*CO1*
32.	a)	Apply a Vector Space Model for document retrieval by creating a term-document matrix for a sample collection.	*10*	*K3*	*CO2*
		OR			
	b)	Develop a method for Relevance Feedback and query expansion to improve search results in an IR system.	*10*	*K3*	*CO2*
33.	a)	Apply a Support Vector Machine classifier for text classification task.	*10*	*K3*	*CO3*
		OR			
	b)	Construct an example of Naive text classification approach for a text data.	*10*	*K3*	*CO3*

34.	a)	Construct Brute force and Knuth-Morris-Pratt algorithm for string matching.	*10*	*K3*	*CO4*
		OR			
	b)	Identify a K-means clustering model and then apply dimensionality reduction before clustering.	*10*	*K3*	*CO4*
35.	a)	Discuss the differences between cluster-based architecture and distributed architecture in the context of search engines.	*10*	*K2*	*CO5*
		OR			
	b)	Explain the various applications of web crawlers in search engines and its architecture.	*10*	*K2*	*CO5*
36.	a)	Explain Content Based Recommendation System in detail.	*10*	*K2*	*CO6*
		OR			
	b)	Discuss the advantages and drawbacks of content-based recommender systems. What limitations do they face in generating diverse recommendations?	*10*	*K2*	*CO6*