Enrolment No./Seat No_____

GUJARAT TECHNOLOGICAL UNIVERSITY

BE - SEMESTER–VII EXAMINATION – SUMMER 2025

Subject Code:3170718 Date:16-05-2025

Subject Name:Information Retrieval

Time:02:30 PM TO 05:00 PM Total Marks:70

Instructions:

1. **Attempt all questions.**
2. **Make suitable assumptions wherever necessary.**
3. **Figures to the right indicate full marks.**
4. **Simple and non-programmable scientific calculators are allowed.**

			MARKS
Q.1	**(a)**	Define an inverted index and explain its role in text retrieval.	**03**
	(b)	What is stemming, and how does it differ from lemmatization?	**04**
	(c)	Explain Structured, Unstructured and Semi-Structured text with suitable example.	**07**
Q.2	**(a)**	List the different applications of Information Retrieval.	**03**
	(b)	Write an algorithm to perform the merging of two posting lists.	**04**
	(c)	Apply the principles of Boolean queries to perform searches on a given document collection using AND, OR, and NOT operators. Take some one line statement in each document to perform queries.	**07**
		OR	
	(c)	Discuss Zipf's Law in the context of text compression and retrieval efficiency. Provide examples of how this law is applied in index compression.	**07**
Q.3	**(a)**	Describe the vector space model in the context of information retrieval. What are the advantages and disadvantages of this model?	**03**
	(b)	Consider the information retrieval system has retrieved 140 positive documents out of which 120 documents were actual positive (correctly retrieved) and 20 documents were actually negative one (wrongly retrieved). Also, 130 negative documents were retrieved out of which 100 documents were actually negative (correctly retrieved) and 30 documents were actually positive one (wrongly retrieved). Find the Precision, Recall, F-measure for positive class and also find the Accuracy of this retrieval system.	**04**
	(c)	Explore the relationship between relevance feedback and query expansion. Discuss their effectiveness in improving retrieval results.	**07**
		OR	
Q.3	**(a)**	What is text classification? Explain its applications in real-world scenarios.	**03**
	(b)	Define: Precision, Recall, F-measure and Accuracy with respect to information retrieval.	**04**

	(c)	Discuss the differences between Naive Bayes, k-Nearest Neighbors, and support vector machines for text classification. Compare their strengths and weaknesses.	**07**
Q.4	**(a)**	Differentiate clustering and classification in the context of text data?	**03**
	(b)	What are kernel functions in SVMs? Explain their role in handling non-linear relationships.	**04**
	(c)	Explain Naive Bayes Classifiers in detail with example.	**07**
		OR	
Q.4	**(a)**	Define hierarchical agglomerative clustering in information retrieval.	**03**
	(b)	Explain the concept of support vector machines (SVMs) for text classification. How do they differ from other classification methods?	**04**
	(c)	Explain in detail the k-means clustering algorithm and discuss its limitations. How can these limitations be mitigated?	**07**
Q.5	**(a)**	What is text summarization? Explain the difference between extractive and abstractive summarization.	**03**
	(b)	How does the HITS algorithm differ from PageRank in evaluating the importance of web pages?	**04**
	(c)	Describe Cross language information retrieval and its limitation in web search.	**07**
		OR	
Q.5	**(a)**	What is question answering? Describe the different types of question answering systems.	**03**
	(b)	Describe the process of ranking web pages using link analysis. How does link analysis improve search results?	**04**
	(c)	Explain the concept of the Semantic Web in detail. How does it use ontologies and linked data to improve search accuracy and data integration?	**07**

****