Enrolment No./Seat No_____

GUJARAT TECHNOLOGICAL UNIVERSITY

BE- SEMESTER–VII EXAMINATION – WINTER 2025

Subject Code:3170718 Date:20-11-2025

Subject Name:Information Retrieval

Time:10:30 AM TO 01:00 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: Structured data, Boolean query, Inverted index	**03**
	(b)	What is Information Retrieval? Explain the process of IR.	**04**
	(c)	Explain text classification using k-nearest neighbor classification approach.	**07**
Q.2	**(a)**	What is token? Which are the different issues in tokenization?	**03**
	(b)	How to compute tf-idf weight for a specific term and document? Explain.	**04**
	(c)	Explain: i) Blocked sort based indexing ii) Dynamic indexing	**07**
		OR	
	(c)	Explain: i) Cosine measure ii) Document length normalization	**07**
Q.3	**(a)**	What is F-measure? Why F-measure is required?	**03**
	(b)	What is precision and recall? An IR system returns 23 relevant documents, and 17 nonrelevant documents. There are a total of 35 relevant documents in the collection. What is the precision and recall of the system on this search?	**04**
	(c)	Discuss about user relevance feedback. Illustrate the cases where use of only relevance feedback is not enough?	**07**
		OR	
Q.3	**(a)**	Identify the requirement of spam filters.	**03**
	(b)	Illustrate boosting technique with any one algorithm.	**04**
	(c)	Explain kappa statistic. How is it interpreted? Consider agreement details of two judges on document relevance. Compute kappa statistic for the given details:	**07**

		Judge 2	
		Relevant	Non-Relevant
Judge 1	Relevant	250	20
	Non-relevant	15	45

Q.4 **(a)** What is text classification? List out examples of text classification in the context of information retrieval. **03**

(b) What is hierarchical agglomerative clustering? Explain. **04**

(c) Discuss: i) N-gram index ii) Zipf's law **07**

OR

Q.4 **(a)** Which are the limitations of k-means clustering? **03**

(b) Explain clustering based on Gaussian mixture model. **04**

(c) Demonstrate SVM classifier. Explain any three kernel functions. **07**

Q.5 **(a)** What do you mean by text summarization? Mention its types. **03**

(b) Explain: Page rank, HITS, XML retrieval, Semantic web **04**

(c) Illustrate cross language information retrieval and the approaches used for it. **07**

OR

Q.5 **(a)** Write short note on question answering. **03**

(b) Discuss web crawler architecture. **04**

(c) Which are the basic tasks in topic detection and tracking? What is story in this context? **07**

****