Enrolment No./Seat No_____

# GUJARAT TECHNOLOGICAL UNIVERSITY

**BE- SEMESTER–VII (NEW) EXAMINATION – WINTER 2024**

**Subject Code:3170718 Date:30-11-2024**

**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, semi-structured and unstructured data. | **03** |
| | **(b)** | Explain kappa measure with an example. | **04** |
| | **(c)** | What is Naïve bayes classifier? Explain text classification using Naïve bayes classifier. | **07** |
| **Q.2** | **(a)** | Differentiate classification and clustering. | **03** |
| | **(b)** | Define tokenization, normalization, stemming and stop words in text processing. | **04** |
| | **(c)** | Write a short note on: i) Zip's law ii) Gamma codes | **07** |
| | | **OR** | |
| | **(c)** | What is tf-idf weighting? Explain tf and idf in detail. | **07** |
| **Q.3** | **(a)** | How cosine similarity measure is used to compare documents? | **03** |
| | **(b)** | Briefly explain k-nearest neighbor classifier. | **04** |
| | **(c)** | Explain Boolean retrieval model and vector space model. | **07** |
| | | **OR** | |
| **Q.3** | **(a)** | Which are the issues with k-means clustering? | **03** |
| | **(b)** | Discuss biword and positional indexes for handling phrase query. | **04** |
| | **(c)** | What is relevance feedback? Explain Rocchio algorithm for relevance feedback. | **07** |
| **Q.4** | **(a)** | Briefly explain spam filtering. | **03** |
| | **(b)** | What is user happiness? How to measure user happiness? | **04** |
| | **(c)** | Explain hierarchical agglomerative clustering with an example. | **07** |
| | | **OR** | |
| **Q.4** | **(a)** | Identify the need of F-measure and describe F-measure in brief. | **03** |
| | **(b)** | Discuss precision and recall. | **04** |
| | **(c)** | Explain support vector machine classifier for text classification. | **07** |
| **Q.5** | **(a)** | What is text summarization? Enlist types of text summarization. | **03** |
| | **(b)** | What is web crawler? How does it work? | **04** |
| | **(c)** | What is ranking in information retrieval? List ranking algorithms and explain any one ranking algorithm. | **07** |
| | | **OR** | |
| **Q.5** | **(a)** | Discuss question answering. | **03** |
| | **(b)** | What is semantic web? What is the need of semantic web? | **04** |
| | **(c)** | Explain primary tasks of topic detection and tracking. | **07** |

**********