*R22 Regulation* *Subject code: 4P6HE*

# TKR COLLEGE OF ENGINEERING AND TECHNOLOGY

**(Autonomous, Accredited by NAAC with 'A+' Grade)**

## B.Tech VI Semester Regular Examinations, May 2025

### INFORMATION RETRIEVAL SYSTEMS

### (CSE(DS))

***Maximum Marks: 60*** Date: 25.06.2025 Duration: 3 hours

**Note:** **1.This question paper contains two parts A and B.**
**2. Part A is compulsory which carries 20 marks. Answer all questions in Part A.**
**3. Part B consists of 5 Units. Answer any one full question from each unit.**
**4. Each question carries 10 marks and may have a, b, c, d as sub questions.**

Part-A

| All the following questions carry equal marks (10x1M=10 Marks) | | Marks | CO | Bloom Tx |
|---|---|---|---|---|
| 1.a | What are the primary objectives of an IR system? | 1M | 1 | L1 |
| b) | What are the different kinds of searching techniques in information retrieval? | 1M | 1 | L1 |
| c) | What are the challenges in XML retrieval? | 1M | 2 | L1 |
| d) | Distinguish between elements and attributes in XML. | 1M | 2 | L2 |
| e) | State any two properties of Naive Bayes classifiers. | 1M | 3 | L1 |
| f) | What are the inputs and outputs of a text classification system? | 1M | 3 | L1 |
| g) | What are the measures used for evaluation of external criterion of quality of clusters? | 1M | 4 | L1 |
| h) | What is the margin in an SVM? | 1M | 4 | L1 |
| i) | What is the role of Web graphs in link analysis? | 1M | 5 | L1 |
| j) | What are the features that web crawlers must and should provide? | 1M | 5 | L1 |

Part-B

| Answer All the following questions. (5X10M=50Marks) | | Marks | CO | Bloom Tx |
|---|---|---|---|---|
| 2 | a) Explain the functional overview of the Information Retrieval System | 5M | 1 | L2 |
| | b) How the Information Retrieval System is related to Database Management System? | 5M | 1 | L3 |
| | OR | | | |
| 3 | a) What is the difference between the concept of a Digital Library and an Information Retrieval System? What new areas of information retrieval research may be important to support a Digital Library? | 5M | 1 | L2 |
| | b) What is a Browse capability? Explain about various browse capabilities. | 5M | 1 | L2 |
| 4 | a) Compare text-centric and data-centric XML retrieval. | 5M | 2 | L4 |
| | b) Explain evaluation of XML retrieval. | 5M | 2 | L2 |
| | OR | | | |
| 5 | a) Explain 1/0 loss case probability ranking principle | 5M | 2 | L2 |

| | | | | |
|---|---|---|---|---|
| | b) Explain XML retrieval with vector space model. | 5M | 2 | L2 |
| 6 | a) Under what conditions would the Bayesian and the Vector approach be the same? | 5M | 3 | L3 |
| | b) Discuss the importance of feature selection in high-dimensional text data. What techniques can be used to select features effectively? | 5M | 3 | L2 |
| | OR | | | |
| 7 | a) Illustrate the Rocchio classification technique and how it updates centroids for different classes. | 5M | 3 | L3 |
| | b) Explain the bias-variance tradeoff with diagrams. How can one balance bias and variance when designing a classifier? | 5M | 3 | L2 |
| 8 | a) Discuss about Hierarchical clustering in Information Retrieval focuses on the area of hierarchical agglomerative clustering methods. | 5M | 4 | L2 |
| | b) Discuss about machine learning methods in adhoc information retrieval. | 5M | 4 | L2 |
| | OR | | | |
| 9 | a) Apply k-means algorithm and find centroids. Take A & B as a initial centroids A(1,2),B(9,10), C(8,6), D(5,5), E(3,3), F(6,1),G(2,2) H(2,8) | 5M | 4 | L3 |
| | b) Will the clustering process always come to the same final set of clusters no matter. What the starting clusters? Explain your answer. | 5M | 4 | L2 |
| 10 | a) What is Web Crawling? Draw and explain the basic Web Crawler architecture. | 5M | 5 | L3 |
| | b) How do we compute page Rank? Explain. | 5M | 5 | L3 |
| | OR | | | |
| 11 | a) Identify some of the key individuals or projects that significantly contributed to the history of web search. | 5M | 5 | L3 |
| | b) How can the Hubs and Authorities algorithm be used to improve the relevance and quality of search results? | 5M | 5 | L3 |