R22 Regulation *Subject code:4P6HE*

TKR COLLEGE OF ENGINEERING AND TECHNOLOGY

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

B.Tech VI Semester Regular/Supplementary Examinations, April 2026

INFORMATION RETRIEVAL SYSTEM

(CSE(DS))

Maximum Marks: 60 Date:23.04.2026 Duration: 3 hours

Note: **1.This question paper contains two parts A and B.**
2. Part A is compulsory which carries 10 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)	Define Information Retrieval System.	1M	CO1	L1
b)	What is indexing in IR systems?	1M	CO1	L1
c)	State challenges in XML retrieval.	1M	CO2	L1
d)	What is meant by text-centric XML?	1M	CO2	L2
e)	Write about KNN.	1M	CO3	L2
f)	State measures used in evaluation of text classification.	1M	CO3	L2
g)	Define linearly separable data.	1M	CO4	L1
h)	What is K-means clustering?	1M	CO4	L1
i)	What do you mean by Hubs and Authorities?	1M	CO5	L2
j)	What is the economic model behind most web search engines?	1M	CO5	L2

Part-B

Answer All the following questions. (5X10M=50Marks)		Marks	CO	Bloom Tx
2	a) Explain the objectives and functional overview of Information Retrieval Systems.	5M	CO1	L1
	b) Compare Information Retrieval Systems with Database Management Systems in at least four points.	5M	CO1	L4
	OR			
3	a) Write a short note on the role of IRS in digital libraries.	5M	CO1	L1
	b) Describe search, browse and miscellaneous capabilities of IRS with suitable examples.	5M	CO1	L4
4	a) Explain basic XML concepts and describe the major challenges in XML retrieval.	5M	CO2	L2
	b) Write a short note on extensions and improvements to the Binary Independence Model in modern IR.	5M	CO2	L4
	OR			

5	a) Explain the Probability Ranking Principle and its importance in probabilistic information retrieval. b) Differentiate between text-centric and data-centric XML retrieval with suitable examples.	5M 5M	CO2 CO2	L4 L4
6	a) Explain the text classification problem in information retrieval and give real-world applications. b) Explain the Bernoulli model for text classification and compare it briefly with the multinomial model.	5M 5M	CO3 CO3	L3 L4
	OR			
7	a) Explain vector space classification and how document representations and similarity measures are used. b) Explain k-nearest neighbour (k-NN) classification for text, highlighting its advantages and drawbacks.	5M 5M	CO3 CO3	L4 L3
8	a) Discuss extensions to the SVM model for non-linearly separable data using kernels. b) Explain the evaluation of clustering quality in information retrieval, mentioning suitable internal or external measures.	5M 5M	CO4 CO4	L4 L3
	OR			
9	a) Explain flat clustering in information retrieval and describe the K-means algorithm in detail. b) Differentiate between single-link and complete-link clustering, mentioning their advantages and disadvantages.	5M 5M	CO4 CO4	L3 L4
10	a) Describe the search user experience and explain how factors like index size, latency, and result quality influence it. b) Explain the concept of near-duplicates and shingling in web search, and why they are important.	5M 5M	CO5 CO5	L5 L5
	OR			
11	a) Describe the architecture and main components of a web crawler, including politeness and freshness issues. b) Derive the basic idea of the PageRank algorithm and discuss the intuition behind its formula.	5M 5M	CO5 CO5	L5 L5