APJ ABDULKALAM TECHNOLOGICAL UNIVERSITY

08 PALAKKAD CLUSTER

Q. P. Code : CSE0821222-I **(Pages: 2)**

Name: .....

Reg. No: .....

SECOND SEMESTER M.TECH. DEGREE EXAMINATION JULY 2021

Branch: Computer Science and Engineering **Specialization: Computer Science and Engineering**

08CS6022 INFORMATION RETRIEVAL

(Common to CSE)

Time: 2 hour 15 minutes **Max. Marks: 60**

Answer all six questions.

Modules 1 to 6: Part 'a' of each question is compulsory and answer either part 'b' or part 'c' of each question.

Q. No.	Module 1	Marks
1. a	Distinguish between Information retrieval and Data retrieval.	3
	Answer b or c	
b	Explain the Retrieval Process on the basis of logical View of Documents.	6
c	i) How the performance of Information retrieval systems is evaluated?	3
	ii) A query-based search on a particular topic, retrieved only 50 records of which 40 are relevant. But the actual database management system contains 80 records on the same topic of which 55 are relevant. With the help of a confusion matrix find the precision and recall scores of the search.	3

Q. No.	Module 2	Marks
2. a	What do you mean by Inverted Index? Give example	3
	Answer b or c	
b	Compare and contrast Boolean model with Vector space model.	6
c	Illustrate the TF- IDF weighting Scheme.	6

Q. No.	Module 3	Marks
3. a	Explain the basic operation of a crawler.	3
	Answer b or c	
b	Explain how a query is formulated according to relevance feedback (Give example).	6
c	Describe briefly how page rank is computed.	6

Q. No.	Module 4	Marks
4. a	Explain the structure of an XML document.	3
	Answer b or c	
b	What are spatial access methods? Explain how spatial access is carried out using R-tree data structure.	6
c	Discuss in detail, the Gemini Algorithm with the help of an example.	6

Q. No.	Module 5	Marks
5. a	Illustrate the role of Decision Tree as a classifier?	4
	Answer b or c	
b	Describe the K-Nearest Neighbour algorithm with the help of an example.	8
c	Explain Naive Bayes Classifier.	3
	Apply Naive Bayes classifier on the data given in table below; Compute P(Yes l x'), P (No l x') for the unseen data x':(Outlook = Sunny, Temperature=cool, Humidity =High, Wind: Strong). Predict the result with proper justification.	5

Day	Outlook	Temperature	Humidity	Wind	Play Cricket
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	High	Weak	Yes
D6	Rain	Cool	Normal	Weak	Yes
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Sunny	Mild	Normal	Strong	Yes
D11	Overcast	Mild	Normal	Strong	Yes
D12	Rain	Cool	Normal	Weak	Yes

Q. No.	Module 6	Marks
6. a	Describe briefly the termination conditions for K-means clustering.	4
	Answer b or c	
b	Explain Hierarchical Agglomerative Clustering. Also differentiate between Single link and Complete link clustering techniques.	8
c	Write the steps of Expectation Maximization Algorithm and explain. Also write notes on Recommender systems.	8