

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST -2 EXAMINATION- 2024

M.Tech-I Semester (CSE)

COURSE CODE (CREDITS):11M1WCI431 (3)

MAX. MARKS: 25

COURSE NAME: Advanced web Mining

COURSE INSTRUCTORS: Dr. Nishant Sharma

MAX. TIME: 1 Hour 30 Minutes

Note: (a) All questions are compulsory.

(b) The candidate is allowed to make Suitable numeric assumptions wherever required for solving problems

Q.No	Question	CO	Marks
Q1	What are the advantages of utilizing HEAD requests in place of GET requests during the process of web crawling? How do these two types of requests differ in terms of their impact on resource usage, efficiency, and data retrieval? In what specific scenarios or circumstances might a web crawler prefer to employ a GET request instead of a HEAD request, and what implications does this choice have for the overall crawling strategy and performance?	CO-1	6
Q2	A text document contains the following sentence: "Web mining is fascinating and constantly evolving." After tokenization and stemming, what would be the resulting terms? Explain how document parsing contributes to effective information retrieval in a web mining context.	CO-2	5
Q3	(a) Describe an abstract model of ranking in information retrieval. Provide an example of how inverted indexes are used in this model. Discuss the advantages of using inverted indexes over traditional indexes. (b) Given a simple directed graph with five web pages (A, B, C, D, E) and the following link structure: $A \rightarrow B$, $A \rightarrow C$, $B \rightarrow C$, $C \rightarrow D$, $D \rightarrow E$, $E \rightarrow A$. Calculate the PageRank of each page after one iteration, assuming an initial PageRank of 1 for each page. Discuss the significance of PageRank in ranking web pages.	CO-2	3+3
Q4	Based on the table given below from GOV2 collection, calculate estimated result set for the query "tropical fish breeding".	CO-3	4

	<i>Word(s)</i>	<i>Document Frequency</i>	<i>Estimated Frequency</i>		
	tropical	120,990			
	fish	1,131,855			
	aquarium	26,480			
	breeding	81,885			
	tropical fish	18,472	5,433		
	tropical aquarium	1,921	127		
	tropical breeding	5,510	393		
	fish aquarium	9,722	1,189		
	fish breeding	36,427	3,677		
	aquarium breeding	1,848	86		
	tropical fish aquarium	1,529	6		
	tropical fish breeding	3,629	18		
Q5	Discuss the importance of stemming as a preprocessing technique for text data. Explain the Porter's Stemmer algorithm, including its key steps, and describe how it helps improve the effectiveness of search engines in retrieving relevant documents. Illustrate your explanation with an example of how the algorithm processes words related to the same root			CO-4	4