Question Paper

Exam Date & Time: 10-Jul-2023 (02:30 PM - 05:30 PM)

Marks: 50

1)

A)

B)

C)

2)



MANIPAL ACADEMY OF HIGHER EDUCATION

Information & Communication Technology Department VI Semester Makeup Examination June-July 2023

INFORMATION RETRIEVAL [ICT 4035]

Duration: 180 mins.

(3)

Section Duration: 180 mins

Descriptive Questions Answer all the questions. Missing data, if any, may be suitable assumed Rank the following documents for a query "shama hit thief" by using the vector space model with log (5) frequency weighting scheme. Doc 1 : Bhama and Shama hit a thief Doc 2: Shama chased a thief with the knife Doc 3: Shama took the knife Doc 4: Bhama afraid of thief Write the output for the quey "president PO/2 obama", "president PO obama", "president POD/2 obama", and "president POD/1 by considering the following documents: Doc 1: president barack Obama Doc 2: president Obama Doc 3: president barack h. obama Doc 4: obama is president Given: A PO/N B: A must appear no more than N terms before B A PO : A must appear immediately before B A POD/N: A must appear no more than N terms before or after B Consider three terms with posting list sizes s1=100, s2=105, and s3=110. For a conjunctive query that includes these 3 terms, whether processing lists in order of size guaranteed to be optimal? Assume the intersection of s1 and s2 has length 100 and the intersection of s1 and s3 length zero Decompose the matrix S by using singular value decomposition (SVD) theorem.

(5)

(2)

 $S = \begin{pmatrix} 4 & 4 \\ -3 & 3 \end{pmatrix}$ A)

- Consider a document with 3 zones with weights g1=0.2, g2=0.31, and g3=0.49 respectively. What B) (3)are all the distinct score values a document may get? Show details of various possible scores.
- (2) C) If in a non-fault -tolerant system with 1000 nodes, each node has 99.9% uptime,

- (i) What is the uptime of the entire system?
- (ii) What is the time (in %) one or more server is down?
- Consider an information need for which there are 6 relevant documents in the collection. Contrast (5) two systems run on this collection by determining the individual average precision for each of the system followed by mean average precision, precision for top 5, recall for top 5, and f-measure for top 5 results. Their top 10 results are judged for relevance as follows (the leftmost item is the top ranked search result):

System 1 R N R N N N N N R R

System 2 N R N N R R R N N N

A)

- B) Determine whether Michum is phonetic alternative of Mitchum or not. Show all the steps (3)
- C) Determine the f1-score by considering an IR system that retrieves a total of 22 documents with a (2) precision of 0.5. Assume, there are a total of 25 documents in the collection.
- 4) Reconstruct the gap sequence and the postings sequences by considering the following gamma (5) coded gaps.
 - (i) 11101101111110010111111111101000111111001

- B) Consider a dictionary with approximately four lakh terms. Estimate the space usage with blocking (3) for block size=4. Illustrate the difference between dictionary search with and without blocking in terms of the required number of comparisons by taking any suitable example.
- C) Illustrate the importance of the zone index by taking any suitable example. (2)
- 5) Determine the transition probability matrix of a random surfer's walk with teleportation (α=0.5) for (5) the graph given below. Find the sequence of probability vectors for 3 iterations by assuming surfer's initial state as node 1.



- B) Identify any three cases/scenarios where relevance feedback alone is not sufficient. (3)
- C) Identify any two issues that needs to be addressed if there is a need to pick a random page from (2) the index of search engine E1 and test whether it is in the index of search engine E2.

