PRESIDENCY UNIVERSITY
BENGALURU

SCHOOL OF ENGINEERING
END TERM EXAMINATION - JUN 2023

Semester : Semester VI - 2020
Course Code : CSE2051
Course Name : Sem VI - CSE2051 - Information Retrieval
Program : ISB

Date : 19-JUN-2023
Time : 9.30AM - 12.30PM
Max Marks : 100
Weightage : 50%

Instructions:
(i) Read all questions carefully and answer accordingly.
(ii) Question paper consists of 3 parts.
(iii) Scientific and non-programmable calculator are permitted.
(iv) Do not write any information on the question paper other than Roll Number.

PART A

ANSWER ALL THE QUESTIONS (5 X 2 = 10M)

1. Define Politeness. What are the two types of politeness with respect to web crawlers?

(CO4) [Knowledge]

2. How does the Latent Semantic Indexing model increase the impact of Probability model.

(CO2) [Knowledge]

3. Which metrics are/is used to measure the proportion of true positive predictions out of all actual positive instances in a classification model?

(CO4)[Knowledge]

4. Define Recomender Systems.

(CO1)Knowledge]

5. What are the impacts of the Neural Network Model in IR?
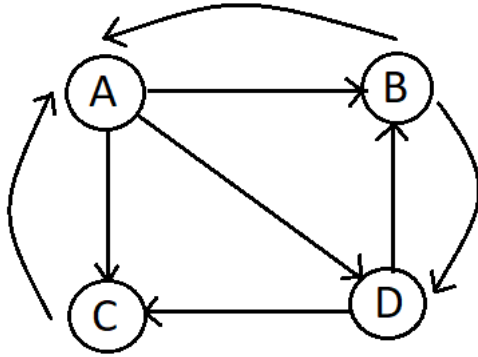
(CO4)Knowledge]

PART B

ANSWER ALL THE QUESTIONS (5 X 10 = 50M)

6. For a set of documents (D1, D2, D3) if term frequency (tf) is 240,150,120 respectively for term (t1) and 130,140,135 respectively for term(t2). Compute the cosine similarity between documents D1, D2 and D3

(CO2) [Comprehension]

7. Find the Hubs and Authorities of all nodes in the following graph

(CO2) [Comprehension]

8. The following list of Rs and Ns represents relevant (R) and nonrelevant (N) returned documents in a ranked list of 20 documents retrieved in response to a query from a collection of 10,000 documents. The top of the ranked list (the document the system thinks is most likely to be relevant) is on the left of the list. This list shows 6 relevant documents. Assume that there are 8 relevant documents in total in the collection.

   R R N N N   N N N R N   R N N N R   N N N N R

   - What is the precision of the system on the top 20 documents?
   - What is the Accuracy of the top 20 documents?

(CO2) [Comprehension]

9. For a query (Q1) there are 3 terms (t1), (t2), and (t3). For the query Q1 there is a document (D1) having 4 terms (t1),(t2),(t3) and (t4). If the document frequency of t1=16, t2=20, t3=30, and t4=60 find the cosine similarity between Q1 and D1.

(CO4) [Comprehension]

10. An IR system returns 8 relevant documents, and 10 nonrelevant documents. There are a total of 20 relevant documents in the collection. What is the precision of the system on this search, and what is its recall?

(CO3) [Comprehension]

PART C

ANSWER ALL THE QUESTIONS   **(2 X 20 = 40M)**

11. The binary independence model (BIM) is used to find the effective weight of a document with respect to the Probability of Relevance and non-relevance as per a given query in two categories. For both categories of the BIM model find the probability of relevance, probability of non-relevance, and EWT for each document given below when N=40, R=8, rt1=3,rt2=4 and rt3=5, Nt1=19, Nt2=14, and Nt3=12.
   DOC1: rhyming swimming and shining
   DOC2: swimming smiling and dialing
   DOC3: styling swimming smiling
   QUERY: rhyming swimming smiling

(CO2) [Application]

12. For the following scenario use the page rank algorithm to find the page rank of all pages:
   In a web graph, 3 pages are present in an instance. Page A has an outdegree to Page B and an in-degree from page B. Similarly, Page B has an outdegree to Page C and an in-degree from page C. In this scenario use N=3 and dumping factor (d)= 0.75 to find the page rank of Page A,B and C.

(CO3) [Application]