CONTINUOUS

Programme Name & Branch: B. Tech

Course Name Code: CSE3024

Course Name: Web Mining

Faculty Name(s): Prof. J. Anuradha, Prof. Shashank Mouli Satapathy & Prof. M. Anbarasi

Class Number(s):1932,1938 &1942 **Exam Duration: 90 mins** **Maximum Marks: 50**

Answer ALL Questions

1. Consider these documents:

Doc 1 The car is driven on the road.
Doc 2 The car and truck are driven on the road
Doc 3 The sky is blue, sun is bright and the truck is blue
Doc 4 The sun is bright in blue sky
Query: blue sky

Find the following

a) TF
b) IDF
c) TFIDF (No need for vector normalization)
d) Using cosine similarity to find the most relevant documents.
e) Using Jaccard similarity to find the most relevant document for the query. (20 M)

2. Explain the web architecture and its security issues. (10 M)

3. a) Decode the following Golomb encoded sequence with b = 10 to detect the sequence of integers passed through the code 00000000000000011101 (5 M)

b) Encode the following three decimal values using Elias Gamma and Elias Delta for 75, 135 and 181 (5 M)

4. a) Draw the static inverted indexing for the following document collection.

Doc 1 The car is driven on the road.
Doc 2 The car and truck are driven on the road
Doc 3 The sky is blue, sun is bright and the truck is blue
Doc 4 The sun is bright in blue sky (5 M)

b) An IR system returns 8 relevant documents, and 10 non relevant documents. There are a total of 20 relevant documents in the collection. What is the precision and recall of the system on this search? (5 M)