USN | | | | | | | | | | |

205MC003

Ph.D./M.S. (Engg.) by Research Degree Examination, April/May 2024
Web Mining

Time: 3 hrs. Max. Marks: 100

Note: ***Answer any FIVE full questions, choosing ONE full question from each module.***

Module-1

1 a. Define Crawling and Indexing. **(04 Marks)**

b. Differentiate between Structured and Unstructured Data Mining. **(06 Marks)**

c. Apply a single thread to implement a web crawler to retrieve web page from www.vtu.ac.in with an algorithm. Explain the steps in detail **(10 Marks)**

OR

2 a. Describe the two common HTTP requesting a page. **(04 Marks)**

b. Differentiate between Clustering and Classification of data. **(06 Marks)**

c. Illustrate Boolean retrival model with examples for the Query operators TRUE or FALSE , Boolean Logic – AND , OR , NOT. **(10 Marks)**

Module-2

3 a. What is Information Retrieval? Give any four applications. Describe the dimensions the field of information retrieval. **(10 Marks)**

b. What is importance of text acquisition in IR? Explain the different aspects of text acquisition in information retrieval. **(10 Marks)**

OR

4 a. Demonstrate the indexing process with a block diagram of a search engine. **(10 Marks)**

b. Examine the process of retrieving a web page using a simple crawler thread algorithm. **(10 Marks)**

Module-3

5 a. Analyse the importance of bottom up and top down approaches in partitioning paradigms with examples. **(10 Marks)**

b. Discuss the need of Preprocessing. With appropriate example discuss the role of data cleaning preprocessing. **(10 Marks)**

OR

6 a. Analyze the importance of Self Organizing Maps and Multi – dimensional Scaling using visualization of IR technique. **(10 Marks)**

b. Discuss with an example of K Means Clustering method. **(10 Marks)**

Module-4

7 a. Discuss the role of clustering a preprocessing tool. Discuss any six requirements of cluster analysis with example. **(10 Marks)**

b. Illustrate the two steps of classification applying for – training data set having attributes (Name, Rank , Years , Experience). **(10 Marks)**

OR

8 a. Apply the decision tree induction algorithm for a training data set having attributes (Age, Income , Student , Credit_rating , Buys_computer) in classifying two classes containing students who buy computer and doesn't buy computer. **(10 Marks)**

b. Apply Navie Bayesian Classifier for a student training data set of having attributes (Name , USN , Sun 1, Sub 2 , Sub 3, GPA) in classifying classes of pass or fail. **(10 Marks)**

Module-5

9 a. Discus the need of page rank algorithm and give the implementation of page rank algorithm with an example and pseudo code. **(10 Marks)**

b. Illustrate the process of retrieving a web page using a simple crawler treat algorithm. **(10 Marks)**

OR

10 a. Two ranking algorithm were implemented over a web search engine and it produced the following results over the query. Evaluate the model using Recall and Precision. **(10 Marks)** Assume 10 relevant documents exists in the system.
System 1 : R R NR R NR
System 2 : NR NR R R R.
Analyse the efficiency of the system by finding the average precision values of the relevant documents and justify which system is better for the search.

b. Illustrate the classification of Monotheist and Polythetic clustering of results. **(10 Marks)**

* * * * *