Vidya Pratishthan's
Kamalnayan Bajaj Institute of Engineering and Technology, Baramati
# Department of Artificial Intelligence and Data Science
Mid-Term Examination
Academic Year 2023-24

Subject:-Elective IV-Information Retrieval                                           Max Marks:-30

Instructions:

1. Question 1 OR Question 2 AND Question 3 OR Question 4

2. Figures to right indicates CO, RBT Level and Marks

3. Use of scientific calculator is allowed

$CO_1$ : Understand the concept of Information Retrieval

1.   1. What are major challenges in IR?                                            $CO_1, L_2$ [5]

   2. Draw the inverted index that would be built for following document collection
      Doc1 : New home sales top forecasts
      Doc2 : Home sales rise in July
      Doc3 : Increase in home sales in July
      Doc4 : July new home sales rise.                                              $CO_1, L_2$ [3]

   3. Discuss Latent Semantic Indexing.                                            $CO_1, L_2$ [7]

OR

2.   1. Explain components of an IR model.                                         $CO_1, L_2$ [5]

   2. Consider following documents:
      Doc1 : breakthrough drug for schizophrenia
      Doc2 : new schizophrenia drug
      Doc3 : new approach for treatment of schizophrenia
      Doc4 : July new hopes for schizophrenia patients
      a) Draw the term document incidence matrix for this document collection
      b) Draw inverted index representation for this collection                    $CO_1, L_2$ [5]

   3. The representation of a set of documents as vectors is in a common vector space. Discuss the
      Information Retrieval Operation for the same.                               $CO_1, L_2$ [5]

   $CO_2$ :Use an indexing approach for retrieval of documents

3.   1. Explain Document at a Time Query Processing.                               $CO_2, L_2$ [4]

   2. Enlist Dictionaries-Types. Discuss Lexicographical type and how it is realised?   $CO_2, L_4$ [6]

   3. Discuss in brief In memory an Sort Based Index Construction                  $CO_2, L_2$ [5]

OR

4.   1. Explain Term at a Time Query Processing.                                   $CO_2, L_2$ [4]

   2. Collisions can be handled by using Hash based dictionary. True or False. Justify your answer
      with proper representation.                                                 $CO_2, L_4$ [6]

   3. Discuss in brief Merge Based and Disk Based Index Construction.              $CO_2, L_2$ [5]

Answer Key: