

Exam Code: 1001  
Sub. Code: 35024

2124  
M.E. (Computer Science and Engineering)  
Third Semester  
CS-8304: Information Retrieval

Time allowed: 3 Hours

Max. Marks: 50

NOTE: Attempt five questions in all, including Question No. 1 which is compulsory and selecting two each questions from Unit.

x-x-x

I. Attempt the following:-

- a) Differentiate among precision, recall, and F1 score metrics.
- b) With one example, explain stem words and positional indexes.
- c) Write a brief note on components of IR.
- d) What is the role of hashing? State its limitations.
- e) Define wild-card searches. Explain the KNN algorithm for classification of data point. (5x2)

**UNIT - I**

- II. a) How stemming and lemmatization are helpful in text analysis? What are the potential implications of using one method over the other?
- b) Compare features of term frequency and inverse document frequency. Why do we use log frequency weighing of term frequency and inverse document frequency? (2x5)

- III. a) Define n-gram indexing and discuss its applications in text retrieval.
- b) What is the difference between dynamic and static indexing? Discuss the challenges and benefits to implement dynamic indexing in real-time applications. (2x5)

- IV. a) Define cosine similarity measure. How is it used to score similarity between documents and queries in the vector space model?
- b) Discuss Uncertainty, Proximity, and Weights in Query Expressions. (2x5)

**UNIT - II**

- V. Discuss Decision Trees and SVM classification techniques. How they handle categorical and continuous data. (10)

P.T.O.



(2)

- VI. a) Explain the role of robots.txt files in web crawling.  
b) Explain the primary components involved in executing a web search, from user queries to search results?  
(2x5)

VII. Write a note on the following:-

- a) Spam filtering
- b) Distributed indexing
- c) Types of indexes
- d) Vector space model

(4x2½)

x-x-x