Exam.Code:1001
Sub. Code: 7639

1078
M.E. (Computer Science and Engineering)
Third Semester
Elective – V
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 questions from each Unit.*

x-x-x

I. Attempt the following:-

a) What is cosine similarity?

b) Differentiate between Recall and Precision.

c) What are the limitations of using hashing for dictionary lookup?

d) Define stemming.

e) What do you understand by spam in the context of web search? (5x2)

UNIT – I

II. a) What is the purpose of using wild card queries? How is permulerm index helpful in handling wild-card queries?

b) Elaborate edit distance technique for correcting spelling errors in queries. (6,4)

III. a) Briefly describe the concept of inverted index and the basic Boolean retrieval model for query processing with the help of an example.

b) Under what circumstances dynamic indexing is preferred and how is it performed? (5,5)

IV. a) How can efficiency of posting lists be enhanced?

b) Discuss the most common approach for processing phrase queries. (5,5)

UNIT – II

V. a) Explain *tf-idf* weighting scheme.

b) Discuss the motivation behind the use of parameter and zone indexes in information retrieval systems. (5,5)

P.T.O.

VI. a) How can we deled and filter near duplicates on the web?

b) How can Naive Bayes Model be used for text classification? (5,5)

VII. Write short note on following:-

a) Vector space scoring

b) Focused web crawler and its architecture (5,5)

x-x-x