Exam.Code:1001
Sub. Code: 7639

1129

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. I which is compulsory and selecting two questions from each Unit.*

x-x-x

I. Attempt the following:-

a) Differentiate between unstructured and semi-structured data.

b) What is the need of token normalization?

c) What kind of index construction is preferred for web search engines?

d) How can you filter email spams using text classification?

e) What is focused web crawler? (5x2)

UNIT – I

II. a) How can you use search trees for dictionary lookup? Also discuss their advantages over hashing in context of dictionary, lookup.

b) Discuss the main idea behind soundex algorithms. (6,4)

III. a) Explain the concept of query optimization taking an example of Boolean queries.

b) Explain the approach used for index construction when the document collections are frequently modified. (2x5)

IV. a) Distinguish between stemming and lemmatization.

b) What do you understand by proximity queries? How can they be handled? (2x5)

UNIT – II

V. a) Discuss k nearest neighbor approach for document classification.

b) Distinguish paid placement strategy with search engine optimization. (2x5)

VI. Discuss the various components of an information retrieval system with a neat sketch. (10)

VII. a) How can you quantify the similarity between a query and a document? Explain the vector space model for ranking documents?

b) How do zone indexes support retrieval of documents? (7,3)

x-x-x