Exam Code: 1001
Sub. Code: 7639

2021
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. Answer the following:-

a) What is an inverted index? What is its utility?
b) How isolated-term correction is different from context-sensitive correction? Discuss in brief.
c) What is the purpose of cosine similarity? Give formula to compute it.
d) What are the advantages and disadvantages of Naive Bayes models? Discuss in brief.
e) What do you understand by *politeness* and *extensible* features of a web crawler? Discuss in brief. (5x2)

UNIT – I

II. a) How the conjunctive queries are processed using an inverted index and the basic Boolean retrieval model? Briefly discus the algorithm for conjunctive queries that returns the set of documents containing each term in the input list of terms. Appropriately assume any required information yourself.

b) What is the goal of stemming and lemmatization? What is Porter's Stemmer? Discuss in brief. (6,4)

III. a) What do you understand by k-gram indexes? How these indexes help in processing wildcard queries? Discuss.

b) What are the basic principles underlying the spelling correction systems? Briefly discuss the algorithm for computing the edit distance between strings s1 and s2. (2x5)

IV. Write notes on the following:-

a) Dynamic Indexing
b) Blocked sort-based indexing (2x5)

P.T.O.

UNIT – II

V. a) What do you understand by the terms document frequency, inverse document frequency and term frequency. How these concepts help in scoring documents? Discuss.

b) What do you understand by vector space model for scoring? Discuss the basic algorithm for computing vector space scores. (2x5)

VI. Draw the diagram for a complete basic search system and explain the necessary components in detail. (10)

VII. a) What do you understand by Decision Trees? How decision trees can be used for classification? Discuss.

b) What do you understand by Link Analysis? Briefly discuss the PageRank Algorithm. (2x5)

x-x-x

x-x-x