1059
B.E. (Computer Science and Engineering)
Eighth Semester
Elective – V
CS-803C: Information Retrieval and Management

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) What are Boolean queries?
- b) Differentiate between unstructured and semi-structured text?
- c) What do you understand by document zone?
- d) What is relevance of Web Crawling with respect to Web Search?
- e) Explain edit distance with an example? (5x2)

UNIT – I

II. a) How index can be optimized using skip lists? Explain with suitable example.

b) What is use of soundex? Explain with an example. (2x5)

III. a) What is tolerant information retrieval? How the wild card queries can be processed?

b) What is inverse document frequency (idf)? (2x5)

IV. a) What is role of stop words in a document? List any five stop words.

b) What so you mean by case folding? Explain with example. (2x5)

UNIT – II

V. Assume the following table represents word frequency of different words for documents Doc1, Doc2, Doc3 and New-Doc. Last line in table shows the category of the documents for Doc1, Doc2 and Doc3. We need to find out the category of New-Doc using k-nearest neighbor algorithm. Assume closeness is defined by Euclidean Distance between two documents.

	Doc 1	Doc 2	Doc 3	Doc 4
Election	10	21	10	23
Chandigarh	21	25	2	12
Delhi	22	11	5	22
Duty	5	8	16	10
Vehicle	4	8	22	10
Category	NORTH	LOCAL	CEO	?

(10)

VI.   Suggest category for New-Doc.
      a)   What is link analysis? How it helps in web search?
      b)   What are mail components of any IR system?   **(2x5)**

VII. a) What is Search Engine Optimization? List any three tools used of SEO.

     b) What is cosine similarity? How it used to calculate similarity between two documents?)   **(2x5)**

x-x-x