Exam Code: 1001
Sub. Code: 7316

2123

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

Time allowed: 3 Hours **Max. Marks: 50**

NOTE: ***Attempt <u>five</u> questions in all, including Question No. 1 (Section-A) which is compulsory and selecting two questions from each Section B-C.***

x-x-x

	Section-A	
1.	a) Differentiate between unstructured and semi-structured text data. b) Discuss role of stop words in text processing and information retrieval. c) Compare positional indices and dynamic indices. d) Discuss the characteristics of Decision Trees and their use in classification. e) Explain concept of nearest neighbor techniques and their relevance in information retrieval.	10
	Section-B	
2.	a) Illustrate with an example how Boolean queries can be applied to retrieve specific information from the database. b) Describe process of lemmatization and how it contributes to improving search accuracy.	5 5
3.	a) Discuss the structure of inverted indices and the basic Boolean Retrieval model. b) Define wild-card query. Provide an example of a wild-card query. c) Describe at least two methods for spelling correction.	5 2 3
4.	a) Define and describe the processes of tokenization, stemming, lemmatization, and stop word removal in text encoding. b) Identify potential bottlenecks in distributed indexing systems and propose strategies to mitigate them.	5 5
	Section-C	
5.	a) Discuss the components of an IR system. b) Why is the odd value of "K" preferred over even values in the KNN Algorithm? c) List down some popular algorithms used for deriving Decision Trees and their attribute selection and attribute discard measures.	3 3 4
6.	a) Explain concept of term weighting in the vector space model and how it influences document ranking. b) Briefly describe how reduced dimensionality approximations and random projection can be used to improve efficiency of vector space scoring.	4 6
7.	a) Explain the process of web crawling and its significance in building web indexes. b) Highlight key differences that make web distinct from other information sources. Discuss both advantages and disadvantages for this source.	6 4

x-x-x