

Exam. Code: 1046

Sub. Code: 6412

2072

M.E. Computer Science and Engineering (Cyber Security)

Second Semester

CSN-8205: 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 Part.

x-x-x

I.	a) Is Boolean information retrieval model same as vector-based information retrieval model? Discuss in brief. (02) b) Are skip pointers useful for queries of the form $x \text{ OR } y$ ? Discuss in brief. (02) c) What is edit distance? What is its utility? Discuss in brief. (02) d) What is slack variable and its role in soft margin SVM classifier? Discuss in brief. (02) e) What is Search Engine Optimization? What is its purpose? Discuss in brief. (02)	
PART I		
II.	a) Explain a typical Information Retrieval Process with the help of a diagram highlighting the steps involved. (05) b) What is tokenization? Discuss any two cases/problems that can make the tokenization process difficult to implement. Also discuss their resolution mechanisms. (05)	
III.	a) How could an IR system combine use of a positional index and use of stop words? What is the potential problem, and how could it be handled? (05) b) What is dynamic indexing? How dynamic indexes can be constructed? Discuss. (05)	
IV.	a) What do you understand by Permuterm index? How does this index help us with wildcard queries? Discuss with the help of a suitable example. (05) b) What are the features of blocked sort-based indexing? Briefly discuss BSBI (Blocked Sort-Based Indexing) algorithm. (05)	
PART II		
V.	a) What is Web Crawling? What are the challenges of searching the Web of linked data? Discuss. (05) b) What is the most commonly used approach to weight the importance of a term in a document? Discuss. (05)	
VI.	a) How does SVM classifier learn linearly separable data? Discuss with the help of a suitable example. (05) b) Briefly discuss Random Projection technique for dimensionality reduction. (05)	
VII.	a) Discuss the steps involved in building a spam filter. Appropriately assume any required information yourself. (05) b) Briefly discuss PageRank algorithm and its working. (05)	

x-x-x