

M.E. 3rd Sem
Dec 2022

Exam Code: 1001
Sub. Code: 7316

2122
M.E. (Computer Science and Engineering)
Third Semester
CS-8304: 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) What are the limitations of Boolean retrieval model? Discuss in brief. (b) How do the terms bitword index and phrase index relate to each other? (c) If $ s_i $ denotes the length of string s_i , can the edit distance between s_1 and s_2 be more than $\max\{ s_1 , s_2 \}$? Discuss in brief. (d) What are the strengths and downsides of Naïve Bayes models? Discuss in brief. (e) What is the formula to compute cosine similarity between two documents d_1 and d_2 ?	(02) (02) (02) (02) (02)
PART I		
II	(a) Consider the following documents: Doc 1 breakthrough drug for schizophrenia Doc 2 new schizophrenia drug Doc 3 new approach for treatment of schizophrenia Doc 4 new hopes for schizophrenia patients i) Draw the term-document incidence matrix for this document collection. ii) Draw the inverted index representation for this collection Appropriately assume any required information yourself. (b) What is a wildcard query? Discuss any one technique to process general wildcard queries.	(05) (05)
III	(a) Briefly discuss the optimization approaches that can be applied to process Boolean queries. (b) What is the goal of stemming and lemmatization? Is stemming same as lemmatization? Briefly discuss any one stemming algorithm.	(05) (05)
IV	(a) We have a two-word query. For one term the postings list consists of the following 16 entries: [4,6,10,12,14,16,18,20,22,32,47,81,120,122,157,180] and for the other it is the one entry postings list: [47] Work out how many comparisons would be done to intersect the two postings lists with the following two strategies. Briefly justify your answers: i. Using standard postings lists ii. Using postings lists stored with skip pointers, with a skip length of \sqrt{P} (b) What do you understand by dynamic indexing? Explain the approach followed in dynamic indexing and discuss its complexity.	(05) (05)
PART II		
V	(a) What do you understand by tf-idf weighting? How is tf-idf weight calculated? Discuss with the help of an example. (b) What do you understand by Decision Tree? How are Decision Trees used for classification? Discuss.	(05) (05)
VI	(a) What is Search Engine Optimization? Briefly discuss the methods that can be used to increase the prominence of a webpage within the search results. (b) What do you understand by Support Vector Machine? How is SVM used for classification? Discuss.	(05) (05)
VII	Write notes on the following: (a) Components of an information retrieval system (b) Near-duplicate detection	(05) (05)

x-x-x