

K. J. Somaiya Institute of Technology, Sion, Mumbai-22
(Autonomous College Affiliated to University of Mumbai)

Nov – Dec 2024	
(B. Tech) Program: Computer Engineering Scheme: III II	
Regular Examination: LY Semester: VII	
Course Code: CEDLC7044 and Course Name: Information Retrieval	
Date of Exam: 23/11/24	Duration: 02.50 Hours Max. Marks: 60

Instructions:

- (1) All questions are compulsory.
- (2) Draw neat diagrams wherever applicable.
- (3) Assume suitable data, if necessary.

Q. No.	Question	Max. Marks	CO	BT level
Q 1	Solve any two questions out of three: (05 marks each)	10		
a)	Define and illustrate precision and recall with examples.		CO3	U
b)	Explain the 'Language Model' & 'Probabilistic Model' in detail.		CO2	U
c)	Why do we require the 'Signature File'? Discuss this using one example.		CO5	U
Q 2	Solve any two questions out of three: (05 marks each)	10		
a)	Using suitable examples, explain the SGML, XML & HTML markup languages.		CO4	U
b)	Build a Suffix Tree for a given below text. Search & examine the patterns in the built Suffix Tree. Justify your answer. {bear, bell, bid, bull, buy, sell, stock, stop}		CO5	An
c)	What is Co-occurrence analysis? Discuss the query refinement technique using this approach.		CO3	U
Q.3	Solve any two questions out of three. (10 marks each)	20		
a)	List the types of Information System and discuss them in detail.		CO1	U
b)	Explain the 'Structure Text Retrieval Model' with suitable examples.		CO2	U
c)	Draw a standard model for the search process and explain in detail about the information access process.		CO6	U
Q.4	Solve any two questions out of three. (10 marks each)	20		
a)	What would the revised query vector be after relevance feedback? Apply the Rocchio relevance feedback algorithm. Given:		CO3	Ap

K. J. Somaiya Institute of Technology, Sion, Mumbai-22
(Autonomous College Affiliated to University of Mumbai)

Nov – Dec 2024	
(B. Tech) Program: Computer Engineering Scheme: II	
Regular Examination: LY Semester: VII	
Course Code: CEDLC7044 and Course Name: <u>Information Retrieval</u>	
Date of Exam: <u>23/11/2024</u>	Duration: 02 30 Hours
Max. Marks: 60	

	● Initial query = "cheap CDs cheap DVDs extremely cheap CDs". ● d1="CDs cheap software cheap CDs" is judged as relevant. ● d2 = "cheap thrills DVDs" is judged as nonrelevant Assume that we are using direct term frequency (with no scaling and no document frequency). There is no need to length-normalize vectors. Assume $\alpha = 1$, $\beta = 0.75$, $\gamma = 0.25$.		
b)	Create and analyze the 'Elias Delta' and 'Elias Gamma' coding and decoding forms separately on the following given numbers a)5 b)3 c)9 d)7	CO4	An
c)	What is the Generic Multimedia Indexing (GEMINI) approach? Apply this technique to a suitable example and discuss the advantages in detail.	CO5	Ap
