

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

Supplementary Exam **Feb-March 2024**

(B.Tech) Program: Computer Engineering Scheme: II

Examination: LY Semester: VII

Course Code: CEDLC7044 and Course Name: Information Retrieval

Date of Exam: _____

28/2/24

Duration: 2.5 Hours

Max. Marks: 60

Instructions:

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

		Max. Marks	CO	BT level
Q 1	Solve any six questions out of eight:	12		
i)	Differentiate between information and data retrieval.	2	CO1	U
ii)	What is the taxonomy of information retrieval models?	2	CO2	U
iii)	What are the main differences between the Boolean model and the Vector Space Model in set-theoretic information retrieval?	2	CO2	U
iv)	What is pattern matching in query languages and how is it used in IR?	2	CO3	U
v)	How does stemming help in document preprocessing?	2	CO4	U
vi)	What does "finding a needle in a haystack" mean?	2	CO5	U
vii)	What are some popular meta search engines?	2	CO5	U
viii)	List some important differences can contribute to acceptance or rejection of interface techniques.	2	CO6	U
Q.2	Solve any four questions out of six.	16		
i)	List and discuss components of the information system.	4	CO1	U
ii)	Compare and contrast between boolean model and Vector Model.	4	CO2	U
iii)	What is a structure query? Explain different types of structure query.	4	CO3	U
iv)	What are the advantages and disadvantages of using text compression in document processing?	4	CO4	U
v)	In what way is the signature approach advantageous over other text retrieval methods?	4	CO5	U
vi)	Elaborate on types of starting points.	4	CO6	U

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

Supplementary Exam: Feb - March 2024
 (B.Tech) Program: Computer Engineering Scheme: II
 Examination: LY Semester: VII
 Course Code: CEDLC7044 and Course Name: Information Retrieval

Date of Exam: *28/2/24* Duration: 2.5 Hours Max. Marks: 60

Q.3	Solve any two questions out of three.	16											
i)	Explain the Retrieval Process with appropriate diagram.	8	CO1	U									
ii)	Illustrate Extended boolean model with advantages and disadvantages	8	CO2	U									
iii)	Create a signature file of the following text by using hash functions . <div style="display: flex; justify-content: space-around; font-weight: bold; font-size: small;"> Block 1 Block 2 Block 3 Block 4 </div> <table border="1" style="width: 100%; border-collapse: collapse; font-size: x-small;"> <tr> <td style="width: 25%;">This is a text.</td> <td style="width: 25%;">A text has many</td> <td style="width: 25%;">words. Words are</td> <td style="width: 25%;">made from letters.</td> </tr> </table> <p>Hash Functions:</p> <table border="1" style="width: 100%; border-collapse: collapse; font-size: x-small;"> <tr><td>h(text) = 000101</td></tr> <tr><td>h(many) = .110000</td></tr> <tr><td>h(words) = 100100</td></tr> <tr><td>h(made) = 001100</td></tr> <tr><td>h(letters) = 100001</td></tr> </table> <p>Search the following patterns in the signature file. a. text b. words c. many d. text has many</p>	This is a text.	A text has many	words. Words are	made from letters.	h(text) = 000101	h(many) = .110000	h(words) = 100100	h(made) = 001100	h(letters) = 100001	8	CO5	AP
This is a text.	A text has many	words. Words are	made from letters.										
h(text) = 000101													
h(many) = .110000													
h(words) = 100100													
h(made) = 001100													
h(letters) = 100001													
Q.4	Solve any two questions out of three.	16											
i)	What is user relevance feedback in query operations? Explain its types in detail.	8	CO3	U									
ii)	Explain with example Document Pre-processing in detail.	8	CO4	U									
iii)	Describe Generic Multimedia Indexing Approach for the Two dimensional color images in detail.	8	CO5	U									
