

Institute of Science and Technology
Bachelor of Science in Computer Science & Information Technology
Model Question

Course Title: Information Retrieval
Full Marks: 60
Course Code: CSC413

Time: 3 hours
Pass Marks: 24
Semester: VII

Group 'A'

Attempt any TWO Questions. (2 × 10 = 20)

1. What is the role of information in human life? How information is retrieved? Explain the architecture. (3 + 7)
2. Define the role of text shingling. Given the following training dataset apply Rocchio algorithm to classify the document "process scheduling". Here the two classes are "operating system" and "Automata". (3 + 7)
operating system :- disk scheduling
operating system :- process management
Automata :- process transition
Automata :- context free grammar
3. How rules are defined in Porter stemmer? Given the following documents and query rank the documents. (4 + 6)
doc1 = "finite state machine"
doc2 = "transition machine"
doc3 = "transition state"
query = "state machine"

Group 'B'

Attempt any EIGHT Questions. (8 × 5 = 40)

4. How inverted index is created? Explain with a suitable example. (5)
5. What is the effect on precision and recall in evaluating ranked documents? Illustrate with an example. (5)
6. Explain the working mechanism of search engine. (5)
7. Why do we need to expand query? How it is performed using relevance feedback? (1 + 4)
8. Differentiate between collaborative and content based recommendation system. (5)
9. How questions and answers are processed in QA system? Explain. (5)
10. Describe the significance of latent semantic indexing and singular value decomposition. (5)
11. How spider works? Write the algorithm. (5)
12. Define stop words. How do you stem the word? (2 + 3)