

Reg. No. : 

--	--	--	--	--	--	--	--	--	--	--	--	--

**Question Paper Code : 70365**

B.E./B.Tech. DEGREE EXAMINATIONS, NOVEMBER/DECEMBER 2021.

Seventh Semester

Computer Science and Engineering

CS 6007 – INFORMATION RETRIEVAL

(Regulations 2013)

Time : Three hours

Maximum : 100 marks

Answer ALL questions.

PART A — (10 × 2 = 20 marks)

1. What is peer-to-peer search?
2. What are the performance measures for search engine?
3. Can the tf-idf weight of a term in a document exceed 1 ? Why?
4. Consider the two texts, “Tom and Jerry are friends” and “Jack and Tom are friends”. What is the cosine similarity for these two texts?
5. What is the purpose of web crawler?
6. What are the requirements of XML information retrieval systems?
7. What is snippet generation?
8. List the characteristics of Map Reduce Strategy.
9. What are the characteristics of information filtering?
10. What are the desirable properties of a clustering algorithm?

PART B — (5 × 13 = 65 marks)

11. (a) List and describe the components of an Information Retrieval (IR) system in detail.

Or

- (b) Identify and describe the components of a search engine with a neat diagram.

12. (a) Briefly explain weighting and cosine similarity.

Or

- (b) Write about relevance feedback and query expansion.

13. (a) (i) Explain in detail about finger print algorithm for near-duplication Detection. (7)

- (ii) Explain the features and architecture of web crawlers. (6)

Or

- (b) (i) Explain about on-line selection in web crawling. (6)

- (ii) Explain in detail about Vector space model for XML retrieval. (7)

14. (a) (i) Explain in detail about Community-based Question Answering system. (8)

- (ii) Brief on Personalized search. (5)

Or

- (b) (i) Explain in detail, the Collaborative Filtering using clustering technique. (8)

- (ii) Brief about HITS algorithm. (5)

15. (a) Write a detailed note on Text mining. (13)

Or

- (b) Explain Agglomerative clustering algorithm. Compare it with other clustering algorithms. (13)

PART C — (1 × 15 = 15 marks)

16. (a) Consider a web graph with three nodes 1,2 and 3. The links are as follows :  $1 \rightarrow 2$ ,  $3 \rightarrow 2$ ,  $2 \rightarrow 1$ ,  $2 \rightarrow 3$ . Write down the transition probability matrices for the surfer's walk with teleporting, for the teleport probability:  $\alpha = 0.5$  and compute the page rank. (15)

Or

- (b) (i) How do the various nodes of a distributed crawler communicate and share URLs? (10)
- (ii) When does relevance feedback work? (5)

---