Reg. No. : ☐☐☐☐☐☐☐☐☐☐☐☐☐

Question Paper Code : 20345

B.E./B.Tech. DEGREE EXAMINATION, NOVEMBER/DECEMBER 2018.

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. Define the term Stemming.

4. Differentiate between relevance feedback and pseudo relevance feedback.

5. What are the applications of web crawlers?

6. Define Search Engine Optimization.

7. What do you mean by item-based collaborative filtering?

8. What is invisible web?

9. What is good clustering?

10. Differentiate information filtering and information Retrieval.

PART B — (5 × 13 = 65 marks)

11. (a) Explain about components of Information Retrieval and Search engine. (13)

Or

(b) (i) Explain the impact of the web on Information Retrieval system. (7)

(ii) Brief about Open source search engine framework. (6)

12. (a) (i) Explain in detail about binary independence model for Probability Ranking Principle (PRP). (7)

(ii) Write short notes on Latent Semantic Indexing (LSI) (6)

Or

(b) (i) How do we process a query using an inverted index and the basic Boolean Retrieval model? (7)

(ii) Describe about how to estimate the query generation probability for query likelihood model. (6)

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 HITS link analysis algorithm. (6)

(ii) Explain in detail about community-based question answering system. (7)

Or

(b) (i) Explain in detail about Collaborative filtering and Content based recommendation system. (7)

(ii) Brief on personalized search. (6)

15. (a) (i) Explain in detail about Naïve bayes classification. (7)

(ii) Explain in detail about Multiple - Bernoulli model. (6)

Or

(b) (i) Explain in detail about k-means algorithm. (7)

(ii) Brief about Expectation maximization algorithm. (6)

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)

———