Reg. No. : | | | | | | | | | | | | |

Question Paper Code : X 20378

B.E./B.Tech. DEGREE EXAMINATIONS, NOVEMBER/DECEMBER 2020
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. Define the term information retrieval. State the functions of information retrieval system.

2. List information retrieval models and some open source search frameworks.

3. Consider a document containing 100 words wherein the word cat appears 3 times. Assume we have 10 million documents and the word cat appears in one thousands of these. Calculate TF-IDF for the same.

4. Describe the differences between vector space relevance feedback and probabilistic relevance feedback.

5. Define meta Crawler.

6. What is index compression ?

7. How can we assign a page rank score to each node of the graph ?

8. What are the Challenges in Cross Lingual Information Retrieval ?

9. Differentiate information filtering from Information Retrieval.

10. What is meant by text preprocessing ?

PART – B **(5×13=65 Marks)**

11. a) Discuss the various component of Information Retrieval with neat sketch and explain how information retrieval process works.

(OR)

b) i) Explain the impact of the web on Information Retrieval system. **(6)**

ii) Discuss the Role of Artificial Intelligence in Information Retrieval. **(7)**

12. a) i) Explain in detail about vector-space retrieval models with an example. **(8)**

ii) Explain about Latent Semantic Indexing method. **(5)**

(OR)

b) i) Explain the Rocchio algorithm for relevance feedback. **(5)**

ii) Discuss briefly about Language model based IR. **(8)**

13. a) i) Illustrate the search engine architecture and processing in detail. **(10)**

ii) Write a detail note on how to measure size of web. **(3)**

(OR)

b) i) Explain briefly about XML retrieval. **(8)**

ii) Write a note on search engine optimization. **(5)**

14. a) i) Explain HITS algorithm of Link Analysis. **(7)**

ii) Write a short notes on Hadoop and Map reduce. **(6)**

(OR)

b) Brief about content based recommendation of documents and products.

15. a) Explain decision tree algorithm with example.

(OR)

b) Explain K-means algorithm of clustering with example.

PART – C **(1×15=15 Marks)**

16. a) i) How can you find similarity between document and query in probabilistic principle using Bayes' rule ? **(7)**

ii) Sort and rank the documents in descending order according to the similarity values :
Suppose we query an IR system for the query "fruits and vegetable sales" The database collection consists of three documents (D = 4) with the following content :

D1 : "Logistics for transportation of fresh vegetables"
D2 : "Price hike on leafy vegetables"
D3 : "Transport of hilly vegetable by helicopters"
D4 : "Fruits are good for health" **(8)**

(OR)

b) Using Naïve Bayes Classifier find the Posterior probabilities of the predictors and response :

	Predictors				**Response**
	Outlook	**Temperature**	**Humidity**	**Wind**	**Class Play=Yes Play=No**
Day1	Sunny	Hot	High	Weak	No
Day2	Sunny	Hot	High	Strong	No
Day3	Overcast	Hot	High	Weak	Yes
Day4	Rain	Mild	High	Weak	Yes
Day5	Rain	Cool	Normal	Weak	Yes
Day6	Rain	Cool	Normal	Strong	No
Day7	Overcast	Cool	Normal	Strong	Yes
Day8	Sunny	Mild	High	Weak	No
Day9	Sunny	Cool	Normal	Weak	Yes
Day10	Rain	Mild	Normal	Weak	Yes
Day11	Sunny	Mild	Normal	Strong	Yes
Day12	Overcast	Mild	High	Strong	Yes
Day13	Overcast	Hot	Normal	Weak	Yes
Day14	Rain	Mild	High	Strong	No