

Nirma University

Institute of Technology

Semester End Examination (IR), December - 2017

B. Tech. in Computer Engineering / Information Technology, Semester-VII
IT702 Information Retrieval Systems

Roll /
Exam No.

Supervisor's Initial
with Date

Time: 3 Hours

Max Marks :100

- Instructions:
1. Attempt all questions.
 2. Figure to right indicate full marks.
 3. Draw neat sketches wherever necessary.
 4. Assume necessary data wherever required, and indicate clearly.

Q.1 Answer the following:

[18]

- [A] Discuss the role of "k" in following: (i) k-means (ii) k-nearest neighbor classifier (iii) k - SVD. [6]
- [B] State with proper justification whether following statement is true or false: "Support Vector Machine is called maximum margin hyperplane classifier." [6]
- [C] Identify any three drawbacks of traditional search engines and give your suggestions on resolving the same. [6]

OR

- [C] How can search engines be fooled to provide desired web pages as top rank documents? Give your suggestions to resolve such issues. [6]

Q.2 Answer the following:

[16]

- [A] Draw architecture of a neural network with backpropagation for classifying an email text as "spam" or "ham". The email is represented by TF/IDF model with 1000 dimensions. How many weight parameters are required? Assume a two-layer neural network with one hidden layer. [8]
- [B] Demonstrate with an example how gradient descent algorithm helps in optimizing the parameters of a neural network. [8]

Q.3 Answer the following:

[16]

- [A] How can one find similarity between two text documents? Illustrate with an example. [6]
- [B] Discuss the cold start problem and suggest methods to address this problem. [6]

OR

- [B] Discuss long tail phenomenon in Retail & Marketing scenario. [6]
- [C] Present your views on the need for a dimensionality reduction technique in an IR system. [4]

Q.4 Do as directed. [18]

- [A] Show with an example how language model is used for next word prediction in a sentence. [6]

OR

- [A] Distinguish between hubs and authorities with proper example. [6]
 [B] Describe the tradeoff between precision and recall for an information retrieval system of your choice using an appropriate example. [6]
 [C] Represent the process of genetic algorithm using a flow chart. [6]

Q.5 Answer the following: [18]

- [A] For the utility matrix shown in Table 1, users have rated the items in the scale of 1 to 5. Compute the following: [12]

Table 1: Utility Matrix for Product ratings

	P1	P2	P3	P4	P5	P6	P7	P8
U1	1	3	4		2		4	
U2		2	3	5	3	3	5	2
U3	2	4	2		4	3		
U4			2	3	5	3	4	5
U5	4	1		3			4	3
U6	4			4	2	3	2	4

- Which user is the most similar to user U4? Verify whether you get the same result with both Jaccard similarity and Cosine similarity. State the reason behind the outcome.
 - What approach would you follow to fill in the blank entries in the utility matrix?
- [B] In the digitization era of today, people have started growing towards cashless and e-commerce based transactions. What do you think would be the role of Information Retrieval systems in this scenario? [6]

Q.6 Answer the following: [14]

- [A] What are issues in building dynamic indexing while merging main and auxiliary indexes? Write and discuss the algorithms of logarithmic merge. [8]
 [B] State the significance of proximity and wildcard operators in Boolean retrieval. [6]

OR

- [B] "The term appearing in a large number of documents in the collection, is probably not significant or discriminative." Give your remarks on this statement with appropriate justification. [6]