Nirma University

Institute of Technology
Semester End Examination (IR), December - 2019
B. Tech. in Information Technology, Semester-VII
IT702 Information Retrieval Systems

Roll /				Supervisor's I	nitio1	
Exam N	No.			with Date		
Time: 3					Max Mar	ks :100
Instruc	tions: 1. Attempt	all questions	•			
	3. Section v	wise separate	licate full mar answer book	to be used.		
	4. Draw ne	at sketches w	herever neces	ssary.		
	5. Assume	necessary da	ta wherever re	equired, and i	ndicate clearly.	
			SECTION -	I		
Q.1 CO1_BL4	representation. D1: "Exam ends today, results awaited." D2: "Admission season will start after exam results." D3: "Enquiry counter is available for admission." Show every step in pre-processing clearly. For the query "admission enquiry", report the ranking order of retrieved					[14]
	documents usi	ng cosine si	milarity.			
Q.2	Can a neural network without any hidden layer implement XOR operation on binary variables? If yes, discuss the process in detail. If no, justify your answer and provide the solution discuss					
[A] CO2_BL3	operation on detail. If no, ju	network with binary varia	ables? If yes	s, discuss t	he process in	[16]
[A] CO2_BL3 [B] CO1_BL4	operation on	network with binary varia stify your ar pare Boolean	ables? If yes nswer and pa n model wit sing an app	s, discuss to rovide the so th vector sp	he process in lution discuss ace model for	
CO2_BL3	operation on detail. If no, ju it in detail. Critically comp	network with binary varial stify your ar pare Boolean esentation u	nswer and particular in model with sing an appropriate or a strategies	s, discuss to rovide the so th vector sp ropriate exar used to ob	he process in plution discuss ace model for mple.	[10]
[B] CO1_BL4	operation on detail. If no, ju it in detail. Critically comp document repr	network with binary varial stify your are pare Boolean esentation under concept and intext of Information	nswer and particular in model with sing an appropriate or a strategies	s, discuss to rovide the so th vector sp ropriate exar used to ob	he process in plution discuss ace model for mple.	[10] [6]
[B] CO1_BL4 [B] CO1_BL2 Q.3 [A]	operation on detail. If no, just in detail. Critically complete document representations of the feedback in confidence of the	network with binary varial stify your are pare Boolean esentation under the concept and intext of Information intext of Information in the concept and in the conce	nswer and particular in model with a sing an appropriate or a strategies in the stra	s, discuss to rovide the so th vector sp ropriate exar used to ob ieval system	he process in plution discuss ace model for mple.	[10] [6] [6]
[B] CO1_BL4 [B] CO1_BL2 Q.3	operation on detail. If no, ju it in detail. Critically comp document reproduction the feedback in contact the feedback in contact the following document reproduction.	network with binary varial stify your are pare Boolean esentation under the concept and intext of Information under the concept and interest and	nswer and particular in model with a sing an appropriate or a strategies in the stra	s, discuss to rovide the so th vector sp ropriate exar used to ob ieval system	he process in plution discuss ace model for mple.	[10] [6]
[B] CO1_BL4 [B] CO1_BL2 Q.3 [A]	operation on detail. If no, ju it in detail. Critically comp document reproduction to the feedback in contact the following document reproduction to the feedback in contact the following document reproduction.	network with binary varial stify your are pare Boolean esentation under the concept and intext of Information in the concept and in the concept and intext of Information in the concept and in th	nswer and particular in model with sing an appropriate of the strategies remation Retricted item rating in the strategies in the strategie	s, discuss to rovide the so the vector spropriate exarused to observed system	he process in plution discuss ace model for mple. tain relevance s.	[10] [6] [6]
[B] CO1_BL4 [B] CO1_BL2 Q.3 [A]	operation on detail. If no, ju it in detail. Critically comp document reproduced by the feedback in contact the feedback in contact the following document reproduced by the feedback in contact the following document reproduced by the feedback in contact the following document reproduced by the feedback in contact the following document for the feedback in contact the following document for the feedback in contact the feedback in conta	network with binary varial stify your are pare Boolean esentation unconcept and ntext of Information unconcept and lowing: 10wing: 11	nswer and property and property and property on the sing an approperty on the strategies of the strate	s, discuss to rovide the south vector spropriate examused to obtieval system. I3 5 1	he process in plution discuss ace model for mple. tain relevance s.	[10] [6] [6]
[B] CO1_BL4 [B] CO1_BL2 Q.3 [A]	operation on detail. If no, ju it in detail. Critically complete document reproduction of the feedback in confeedback in confe	network with binary varial stify your are pare Boolean esentation under the concept and intext of Information under the concept and interest and	nswer and property on model with sing an appropriate of the strategies remation Retrestives are retrestive of the strategies of the strate	s, discuss to rovide the south vector spropriate examused to observed system matrix.	he process in plution discuss ace model for mple. tain relevance s.	[10] [6] [6]
[B] CO1_BL4 [B] CO1_BL2 Q.3 [A]	operation on detail. If no, ju it in detail. Critically comp document reproduced by the feedback in contact the feedback in contact the following document reproduced by the feedback in contact the following document reproduced by the feedback in contact the following document reproduced by the feedback in contact the following document for the feedback in contact the following document for the feedback in contact the feedback in conta	network with binary varial stify your are pare Boolean esentation unconcept and ntext of Information unconcept and lowing: 10wing: 11	nswer and particular	s, discuss to rovide the south vector spropriate examused to obtieval system. I3 5 1	he process in plution discuss ace model for mple. tain relevance s.	[10] [6] [6]

Please note that the empty cell in the matrix denotes that the item is yet not rated. Use user-based collaborative filtering to estimate the user U1's rating for the item I2. Use Pearson correlation to calculate similarity between users. Use the

following formula to compute the rating
$$P_{a,i} = \overline{r}_a + \frac{\sum_{u \in U} (r_{u,i} - \overline{r}_u) \cdot w_{a,u}}{\sum_{u \in U} |w_{a,u}|},$$

 $P_{a,i}$ is the rating of user a for the item i. $\overline{r_a}$ is the average rating of user a. U is the set of 3 users most similar to user a and who have rated the item i. w_{au} is the similarity between users a and u.

Following documents are represented by Term Frequency (TF) [B][8] weight vector. Apply k means algorithm on these documents to partition them into two clusters. Run the algorithm for two iterations.

OR

[B]Discuss about the long tail phenomenon which is typically observed in Retail & Marketing scenario.

SECTION - II

Q.4 Do as directed.

[20]

[A] Can information retrieval be useful in medical domain? Justify [6] CO3_BL2 your answer.

Consider a meta search system with five underlying search [14] [B]CO2_BL4 system, which have ranked four candidate documents or pages a, b, c, and d as follows:

System 1: a, b, c, d

System 2: b, a, d, c System 3: c, b, a, d

System 4: c, b, d

System 5: c, b

Use Borda, Condorcet, and Reciprocal ranking methods to compute the final ranking of the meta search system.

Q.5 Answer the following:

[**18**]

[A] For the following corpus, apply naive Bayesian classification for spam email detection. Assume that the documents are already pre-processed. Fit multinomial distribution to the data.

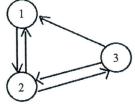
Text	Class
Travel offer booking discount	spam
Offer university graduation	non-spam
Booking offer	spam
Graduation travel	non-spam

Use following text as test sample: university travel.

[B] Given a choice, which system would you prefer: the one which [6] can handle (a) Boolean queries, or (b) phrase queries, or (c) proximity queries, or (d) natural language queries? Comment on each.

Q.6 Consider the following web graph.

[12]



Calculate PageRank of each of the pages in the web graph using 4 iterations of power iteration method. Assume damping factor d = 0.8.

OR

Q.6
CO3_BL4
Assume a text corpus of 100 documents. Documents in this corpus are belonging to 2 categories namely cricket and movies. Assume that documents are represented as TFIDF vectors of size 200. If we wish to use genetic algorithm to classify these documents in two classes, suggest suitable chromosome encoding, fitness function selection method, crossover operator and mutation operator.