Nirma University

Institute of Technology

Semester End Examination (IR/RPR), December - 2018

B. Tech. in 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. Figures to right indicate full marks.
3. Draw neat sketches wherever necessary.
4. Assume suitable assumptions and specify them.

Q-1. Answer the following: [16]

A) Draw and explain the architecture of Information Retrieval systems in detail. 08

B) Compute the eigen values and eigen vectors of following matrix: 08

$$A = \begin{bmatrix} 2 & 1 & 1 \\ 3 & 2 & 1 \\ 2 & 1 & 2 \end{bmatrix}$$

OR

B) Execute k means clustering algorithm on following dataset and show two cluster allocations formed after the second iteration of the algorithm. 08
A(4,6) B(2,5) C(9,3) D(6,9) E(7,5) F(5,7)

Q-2. Answer the following: [16]

A) Which IR technique would you use for implementing Image Captioning systems? Discuss. 04

B) Which kind of data structures are most suitable for storing retrieval models? Explain in detail. 05

C) What are the consequences of avoiding stemming in text processing? Present a scenario to describe your thoughts. 04

D) How does the visualization takes place in IR systems? Show with an example of generating a word cloud from a text corpus. 03

OR

D) How hubs and authorities work during Web search? Elaborate the process with an example. 03

Q-3. Answer the following: [18]

A) With an example, show how proximity based queries are more significant than boolean queries. 06

B) Consider following four documents: 12
D1: Welcome to this university.
D2: All students are welcome here in the campus.
D3: The campus is very pleasant.
D4: How did you welcome students at the campus?

1. Construct term-document matrix using boolean retrieval model. 3
2. For the query "student welcome", which documents will be returned by this model? Illustrate the process. 2
3. Construct the TF-IDF representation of the term document matrix using vector space model. 5
4. Generate an inverted index mechanism for this corpus. 2

Q-4. Answer the following: [16]

A) Show using a diagram the architecture of a multilayer perceptron with two hidden layers. The network is built to classify images of cats and dogs with resolution of 32x32. 08
   1. How many neurons are there in input layer?
   2. How many neurons are there in output layer?
   3. Determine the number of weight parameters for the network.

OR

A) Discuss the role of "gradient descent" in ANN. How does it help in parameter optimization? 08

B) Elaborate the concept of n-grams in language models for Information Retrieval. 04

C) The naive Bayesian classification follows the class conditional assumption. Describe this concept in detail. 04

Q-5. Answer the following: [16]

A) What do you think is the ideal value of "k" in k-nearest neighbor algorithm? Support your answer with concrete example. 05

B) Compute the similarity between following two vectors using Cosine similarity. 06
   V1<1,0,0,1,1,0>
   V2<1,1,0,0,1,0>
   Are these vectors orthonormal?

C) What is search personalization? How does it affect users of online market? 05

OR

C) How the page rank of any webpage is calculated by search engines? Illustrate the process. 05

Q-6. Answer the following: [18]

A) How can IR help in content based audio retrieval? Discuss. 06

B) How machine learning is related to Information Retrieval? Justify. 06

C) Distinguish between precision and recall measures for evaluating IR systems by taking a suitable example. 06