Nirma University

Institute of Technology

Semester End Examination (RPR), May - 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) Represent the concept of Text Processing using flow chart and explain each component with appropriate example. 08

B) Illustrate the role of Singular Value Decomposition in Dimensionality reduction using an example. 08

OR

B) Consider a postings intersection between this posting list with skip pointers (consider skip length $\sqrt{p}$, where p is the length of the posting list.) 08

2 4 5 8 9 11 15 18 19 22 45 49 55 58 61 64 68 69 72 80 85 90 95 99 111 135

And the following intermediate result posting list (which hence has no skip pointers):

2 11 45 62 72 85 97

Trace through the postings intersection algorithm.

1. How often a skip pointer is followed?
2. How many postings comparisons are made by this algorithm while intersecting these two lists?
3. How many postings comparisons would be made if the postings lists are intersected without the use of skip pointers?

Q-2. Answer the following: [16]

A) Which IR technique would you use for Question Answering systems? Discuss. 04

B) How is the evaluation for information retrieval systems carried out? Explain with appropriate example. 05

C) Are numbers/numeric values important and should they be removed or kept during stop word removal? Discuss. 04

D) What is the idf value of a term that occurs in every document? Discuss. 03

OR

D) Can the tf-idf weight of a term in a document exceed 1? Explain. 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 three documents: 12

D1: This is final exam.

D2: Finally the day has arrived.
D3: Exams arrive as soon as nothing.

1. Construct term-document matrix using boolean retrieval model. 3
2. For the query "exam day", 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 neural network with two hidden layers. The network is built to classify MNIST DIGITS data. 08
   1. How many nodes are there in input layer?
   2. How many nodes are there in output layer?
   3. Determine the number of weight parameters for the network.

OR

A) Discuss the role of "learning rate" in ANN. How does it affect the gradient descent algorithm? Describe using an appropriate example. 08

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

C) What is Laplacian correction in context of Bayesian Classification? 04

Q-5. Answer the following: [16]

A) State with justification why kNN is considered as a lazy learner. 05

B) Compute the similarity between following two vectors using (i) Euclidean distance (ii) Cosine similarity (iii) Jaccard similarity. 06
   V1<1,0,0,1,1,0>
   V2<1,1,0,0,1,0>

C) Discuss the cold start problem and strategies to resolve it in Recommender Systems. 05

OR

C) Discuss the long tail phenomenun in context of Recommender Systems. 05

Q-6. Answer the following: [18]

A) With an example, show how Information Retrieval takes place from multimedia documents. 06

B) Distinguish between content based and collaborative filtering in context of Movie Recommender Systems. 06

C) How do search engines present search results to the user? Depict the entire process using a flow chart. 06