Code No: 815BJ **R13**

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD
MCA V Semester Examinations, January - 2018
INFORMATION RETRIEVAL SYSTEMS

Time: 3 Hours **Max. Marks: 60**

Note: This question paper contains two parts A and B.
Part A is compulsory which carries 20 marks. Answer all questions in Part A. Part B consists of 5 Units. Answer any one full question from each unit. Each question carries 8 marks and may have a, b, c as sub questions.

PART - A

5 × 4 Marks = 20

1.a) How to obtain the character sequence in a document? Explain. [4]
b) How to represent query as vectors? Explain. [4]
c) Write down the equations for chain rule, partition rule and Bayes' rule. [4]
d) What are multiclass SVMs? Explain. [4]
e) What are the features that web crawlers must and should provide? Explain. [4]

PART - B

5 × 8 Marks = 40

2. Where wildcard queries are used? Explain about general wildcard queries and k-gram indexes for wildcard queries. [8]

OR

3. What do you know about distributed indexing? Explain with an example of distributed indexing with MapReduce. [8]

4. Explain about
a) Static quality scores and ordering b) Cluster pruning. [4+4]

OR

5.a) Describe the process of evaluating information retrieval system.
b) Discuss about Rocchio algorithm for relevance feedback. [4+4]

6. Write short notes on:
a) Finite automata and language models b) Types of language models. [4+4]

OR

7. Discuss the following text classification models
a) Multinomial unigram language model b) Bernoulli model [4+4]

8. Define hierarchical clustering. Explain about hierarchical agglomerative clustering.[8]

OR

9. Write down the matrix diagonalization theorem. Explain about latent semantic indexing.[8]

10. Discuss the web characteristics in detail. [8]

OR

11. Explain about Markov chains process in detail. [8]

---oo0oo---