

Code No: 815BJ

**JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD****MCA V Semester Examinations, December - 2019****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) What do you understand by the terms “grepping” and “incidence matrix”? Explain.[4]
- b) Explain about weighted zone scoring. [4]
- c) What do you understand by text-centric and data-centric XML retrieval? Explain. [4]
- d) What do you understand about support vector machines? Explain. [4]
- e) What is a web graph? Explain. [4]

**PART - B****5 × 8 Marks = 40**

2. Explain the process of querying using an inverted index and the basic boolean retrieval model. [8]

**OR**

3. What do you understand by tokenization? Explain in detail. [8]

4. Explain about term frequency and weighting in detail. [8]

**OR**

5. Explain the following components of an information retrieval system  
a) Tiered indexes      b) Query-term proximity. [4+4]

6. Discuss the vector space model for XML retrieval. [8]

**OR**

7. Explain about probability ranking principle. [8]

8. Explain the hierarchical clustering in information retrieval. [8]

**OR**

9. Discuss the k-means flat clustering algorithm. [8]

10. Discuss about crawler architecture in detail. [8]

**OR**

11. Write short notes on:  
a) Spammers  
b) Informational, navigational and transactional queries. [4+4]

---ooOoo---