

Code No: 117DX

R13

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD

B. Tech IV Year I Semester Examinations, March - 2017

INFORMATION RETRIEVAL SYSTEMS

(Common to CSE, IT)

Time: 3 Hours

Max. Marks: 75

Note: This question paper contains two parts A and B.

Part A is compulsory which carries 25 marks. Answer all questions in Part A.

Part B consists of 5 Units. Answer any one full question from each unit.

Each question carries 10 marks and may have a, b, c as sub questions.

Part- A

(25 Marks)

- 1.a) What is a non-binary independence model? [2]
- b) What is a term frequency and normalized term frequency? Write down their equations. [3]
- c) Give an example that improves the effectiveness of Information retrieval system. [2]
- d) What is Ward's method in clustering? Give example. [3]
- e) What are semantic networks? [2]
- f) What is comparable corpus and parallel corpus? [3]
- g) What is meant by query processing? [2]
- h) What is a signature and how to construct signature file. [3]
- i) What is high-precision search? [2]
- j) What is structured data and what is the use of XML? [3]

Part-B

(50 Marks)

2. Explain about vector space model in detail. [10]

OR

- 3.a) Explain about retrieval strategies and their categories.

- b) What is smoothing in language model? Explain. [5+5]

- 4.a) Explain how Thesaurus are used to expand a query.

- b) Explain about the use of manually generated Thesauri. [5+5]

OR

5. Explain about:

- a) Resultset clustering
- b) Hierarchical Agglomerative clustering

- c) Rocchio clustering [3+4+3]

- 6.a) What are the four core questions to cross the language barrier?

- b) Explain about document translations and query translations. [4+6]

OR

7. Explain the following in semantic networks

- a) R-distance
- b) K-distance [5+5]

8. Discuss about Duplicate document detection. [10]

OR