

Code No: B2510

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD
M.TECH II SEMESTER EXAMINATIONS, APRIL/MAY 2012
INFORMATION RETRIEVAL SYSTEMS
(SOFTWARE ENGINEERING)

Time: 3hours

Max.Marks:60

Answer any five questions
All questions carry equal marks

- - -

- 1 a) What is Information Retrieval System? Explain about *Functional overview* of Information Retrieval Systems.
b) What is the difference between the concept of a *Digital Library* and an *Information Retrieval*?
- 2.a) Describe the rationale why use of *proximity* will improve *precision* versus use of just the Boolean functions. Discuss its effect on improvement of *recall*.
b) What is the relationship between *vocabulary browse* and *thesauri/concept classes*?
- 3.a) How does the process of *information extraction* differ from the process of *document indexing*?
b) Explain with an example, how *Automatic indexing* process is handled in IRS.
- 4.a) What is *stemming* process? Apply the *Porter stemming* steps to the following words: *irreplaceable, informative, activation and triplicate*.
b) What is *PAT*? How items are represented using it.
- 5.a) What are the benefits of a *weighted index* system over a *Binary index* system?
b) Describe how use of *Natural Language Processing* will assist in the disambiguation process. What is the impact on index structure and the user search interface to take advantage of the results of disambiguation?
- 6.a) Describe the need for *information visualization*. Under what circumstances is it not useful?
b) Is the use of *positive feedback* always better than using *negative feedback* to improve a query?
7. Use the *Boyer-Moore* text search algorithm to search for the term *FANCY* in the text string "*FANCIFUL FANNY FRUIT FILLED MY FANCY*".
 - a. Show all of the steps and explain each of the required character shifts.
 - b. How many character comparisons are required to obtain a match?

8. Given the following Term-Term matrix:

	T1	T2	T3	T4	T5	T6	T7	T8	T9
T1		14	9	0	3	0	12	0	16
T2	14		0	6	4	0	14	0	11
T3	9	0		12	7	4	1	0	14
T4	0	6	12		3	0	14	9	8
T5	3	4	7	3		12	6	16	0
T6	0	0	4	0	12		9	2	9
T7	12	14	1	14	6	9		0	12
T8	0	0	0	9	16	2	0		8
T9	16	11	14	8	0	9	12	8	

- Determine the *Term Relationship* matrix using a threshold of 10 or higher
- Determine the clusters using the *clique technique*.
- Determine the clusters using the *single link technique*.
