

Code No: 07A71201

R07

Set No. 2

IV B.Tech I Semester Examinations, May 2011
INFORMATION RETRIEVAL SYSTEMS
Information Technology

Time: 3 hours

Max Marks: 80

Answer any FIVE Questions
All Questions carry equal marks

1. What are the problems that need resolution when the concept of dissemination profiles expands to including existing data structures? [16]
2. Explain how the stemming algorithm removes suffixes and prefixes? [16]
3. (a) What are the objectives of indexing?
(b) Explain about manual indexing? [8+8]
4. Explain complete term relation method? Give an example of a database with 5 items and 8 terms? Explain? [16]
5. (a) What happens when the weight is equal to the term frequency?
(b) What are the benefits of a weighted index system over a binary index system?
(c) Are there benefits that the binary system can provide over a weighted system? [6+5+5]
6. What algorithmic basis is used for the GESCAN and Fast Data Finder hardware text search machines? Why was this approach used over others? [16]
7. Why should researches in information retrieval care about standards? [16]
8. Explain the Jaccard and the Dice similarity measures? [16]

Code No: 07A71201

R07

Set No. 4

IV B.Tech I Semester Examinations, May 2011
INFORMATION RETRIEVAL SYSTEMS
Information Technology

Time: 3 hours

Max Marks: 80

Answer any FIVE Questions
All Questions carry equal marks

1. (a) Compare clustering with existing clusters for both Term and Item models?
(b) Explain the centroids for Items? Explain item clustering with initial clusters? [4+12]
2. What is the relationship between vocabulary browse and thesauri/concept classes? [16]
3. What does the Selective Dissemination of Information Process provide? [16]
4. Explain RetrievalWare system's approach to queries and ranking? [16]
5. (a) Explain about Public File indexer?
(b) What is a weighted indexing process?
(c) Explain probabilistic model? [5+6+5]
6. Explain the first use of n-grams? How n-grams is used in spelling error detection and correction? [16]
7. (a) How can natural language processing reduce errors in determining phrases?
(b) Explains the steps in a natural language to determine phrases? [6+10]
8. Give the reasons for the interest in the growth of information retrieval systems? [16]

Code No: 07A71201

R07**Set No. 1**

IV B.Tech I Semester Examinations, May 2011
INFORMATION RETRIEVAL SYSTEMS
Information Technology

Time: 3 hours**Max Marks: 80**

Answer any FIVE Questions
All Questions carry equal marks

1. Explain about iterative search and search history log? [16]
2. Define inverted file structure? Explain the files in inverted file structures? [16]
3. Compare the three learning algorithms used for the classification process? [16]
4. Explain the rank-frequency law of Zipf? Represent the distribution of words in TREC database? [16]
5. Explain the advantages of probabilistic approach? Give an example where the probabilistic approach may be applied? [16]
6. Explain about the history of indexing? [16]
7. (a) What is polythetic clustering?
(b) Prove that a term cannot be found in multiple clusters when using the single link technique? [6+10]
8. 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+8]

Code No: 07A71201

R07

Set No. 3

IV B.Tech I Semester Examinations, May 2011
INFORMATION RETRIEVAL SYSTEMS
Information Technology

Time: 3 hours

Max Marks: 80

Answer any FIVE Questions
All Questions carry equal marks

1. Explain the guidelines on the characteristics of the classes? [16]
2. (a) What is browse capability? Explain the possible ways to display a summary of items that are associated with a query?
(b) Explain the process of zoning during display of an item? [8+8]
3. (a) Define vector space model? Represent a three dimensional vector assuming three processing tokens: Petroleum, Mexico, and Oil?
(b) What is the use of audio transcription? [8+8]
4. Explain the Hypertext history? [16]
5. (a) Explain the two classes of index files?
(b) What is the purpose of a structured Database Management System? [8+8]
6. (a) Does high specificity always imply high exhaustivity? Justify your answer?
(b) Will low exhaustivity effect precision and recall? If so, how? [8+8]
7. Describe how other senses could be used in displaying results from searches? [16]
8. Explain about query processor? Explain GESCAN text array processor? [16]
