Code No: R05411201 **Set No. 1**

IV B.Tech I Semester Regular Examinations, November 2008
INFORMATION RETRIEVAL SYSTEMS
(Information Technology)

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

Answer any FIVE Questions
All Questions carry equal marks

⋆ ⋆ ⋆ ⋆ ⋆

1. (a) How the Information Retrieval System is related to Database Management System.
   (b) Explain various types of Information Retrieval Systems. [6+10]
2. Explain the following Browse capabilities:-
   (a) Vocabulary
   (b) Ranking
   (c) Canned query
   (d) Multimedia. [4×4=16]
3. (a) Differentiate Full Item indexing, Public File Indexing and Private File Indexing.
   (b) What is a precoordination? How it is different from linkages? [10+6]
4. (a) Explain in detail about stemming process.
   (b) Describe signature File structure. [8+8]
5. (a) Explain about classes of Automatic Indexing.
   (b) With a neat diagram explain the data flow in an information processing system. [8+8]
6. Differentiate between Manual Clustering and Automatic Term Clustering. Explain with suitable examples. [16]
7. (a) What are the six key characteristics of intelligent agents?
   (b) What are the processing steps used in automatic relevance feed back to enhance user query? [6+10]
8. Clearly explain the following algorithms with suitable examples:
   (a) Boyer-Moore text search algorithm.
   (b) Knuth-Pratt-Morris algorithm.
   (c) Aho-Corasick algorithm. [6+5+5]

⋆ ⋆ ⋆ ⋆ ⋆

Code No: R05411201 **Set No. 2**

IV B.Tech I Semester Regular Examinations, November 2008
INFORMATION RETRIEVAL SYSTEMS
(Information Technology)

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

Answer any FIVE Questions
All Questions carry equal marks

⋆ ⋆ ⋆ ⋆ ⋆

1. Write short notes on:
   (a) Private Index file
   (b) Digital library
   (c) Data warehouse
   (d) Information Retrieval System. [4×4=16]
2. What is a Browse capability? Explain about various browse capabilities. [16]
3. Write short notes on:
   (a) Multimedia indexing
   (b) Indexing by Term
   (c) Indexing by concept. [5+5+6]
4. (a) Describe how the PAT Data structure is different from n-grams.
   (b) Define a S1 string. Explain it with an example. [8+8]
5. Explain in detail about Probabilistic Weighting with an example. [16]
6. Write short notes on the following with suitable examples:
   (a) Manual clustering.
   (b) Automatic Term clustering. [16]
7. (a) Describe the need for information visualization.
   (b) Under what circumstances is information visualization is not useful? Quote an example. [6+10]
8. List out the differences between Boyer-Moore Algorithm and Knuth-Pratt algorithm. Explain. [16]

⋆ ⋆ ⋆ ⋆ ⋆

Code No: R05411201 **Set No. 3**

IV B.Tech I Semester Regular Examinations, November 2008
INFORMATION RETRIEVAL SYSTEMS
(Information Technology)

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

Answer any FIVE Questions
All Questions carry equal marks

⋆ ⋆ ⋆ ⋆ ⋆

1. Explain the following:-
   (a) Selective Dissemination of Information
   (b) Index Data search
   (c) Document database search. [5+6+5]
2. Write short notes on:
   (a) Ranking
   (b) Thesaurus
   (c) Fuzzy searches
   (d) Proximity. [4×4=16]
3. (a) Explain the concept of catalogs.
   (b) What is information extraction? Explain its concepts. [8+8]
4. Write short notes on:
   (a) Inverted File structures.
   (b) N-grams.
   (c) Dictionary look-up stemmers. [6+5+5]
5. What are the problems with weighting schemes and vector model? [16]
6. List out the various techniques in Automatic Term Clustering. Explain. [16]
7. (a) What are the set of rules postulated by Gestalt psychologists for visualization.
   (b) What are the main aspects of human visualization process? Explain. [6+10]
8. (a) Write short notes on:
      i. Precision
      ii. TURR.
   (b) Discuss in detail GESCAN Text Array Processor. [8+8]

⋆ ⋆ ⋆ ⋆ ⋆

Code No: R05411201 **Set No. 4**

IV B.Tech I Semester Regular Examinations, November 2008
INFORMATION RETRIEVAL SYSTEMS
(Information Technology)

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

Answer any FIVE Questions
All Questions carry equal marks

⋆ ⋆ ⋆ ⋆ ⋆

1. Explain the following concepts:-
   (a) Precision & Recall
   (b) Information Retrieval System
   (c) Private Index File
   (d) Push and pull process.

   [4×4=16]

2. (a) What is a Thesaurus? Explain in detail semantic thesaurus.
   (b) Discuss about:
      i. Natural language queries
      ii. Multimedia queries. [8+4+4]

3. (a) Discuss about different mechanisms used in the process of Multimedia indexing.
   (b) Differentiate between precoordination and postcoordination. [8+8]

4. Discuss the following concepts:
   (a) Hyper Text and XML
   (b) Signature File Structure
   (c) Porter stemming algorithm. [5+5+6]

5. What is concept indexing approach? Explain in detail with a suitable example.[16]

6. Distinguish term clustering and item clustering. Discuss suitable examples. [16]

7. List out the Information visualization technologies. Explain each one, with examples. [16]

8. Consider the following table of relevant items in ranked order from four algorithms along with the actual relevance of each item. Assume all algorithms have the highest to lowest relevance is from left to right (Document 1 to last item). A value of Zero implies the document was non-relevant.

Document	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Algo 1	1	0	0	1	1	1	0	0	1	1	0	0	1	1
Algo 2	0	1	1	0	1	1	1	0	0	1	1	0	1	1
Algo 3	0	1	0	0	1	1	1	1	1	0	1	1	1	1
Actual	1	1	1	1	0	0	1	1	1	0	0	1	1	1

Document	15	16	17	18	19	20	21	22	23	24	25	26	27
Algo 1	1	0	0	1	1	1	0	0	1	1	0	0	1
Algo 2	0	1	1	0	1	1	1	0	0	1	1	0	1
Algo 3	0	1	0	0	1	1	1	1	1	0	1	1	1
Actual	1	1	1	1	0	0	1	1	1	1	0	1	1

(a) Calculate and graph precision/recall for all the algorithms on one graph.

(b) Calculate and graph fallout/recall for all the algorithms on one graph.

(c) Calculate the TNRR and TURR for each algorithm (assume uniquely found is only when one algorithm found a relevant item)

(d) Identify which algorithm is best and why? [4×4]

⋆ ⋆ ⋆ ⋆ ⋆