

Code No: 07A71201

R07

Set No. 2

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

Time: 3 hours

Max Marks: 80

Answer any FIVE Questions
All Questions carry equal marks

1. (a) Explain about vectorized information systems.
(b) What is a Bayesian network? Explain two level Bayesian network. [4+12]
2. (a) Explain the goals of Text retrieval conference.
(b) What are the two types of retrieval examined at TREC?
(c) What is the purpose of pooling method? [5+6+5]
3. Explain the steps in manual clustering process with an example. [16]
4. Explain the problems with the vector model. [16]
5. Explain the differences between Information Retrieval Systems and DBMS. [16]
6. Discuss the difficulties of a user being able to correlate his search to the Hit file.
What approach is to be used to overcome these problems? [16]
7. (a) For the word “duplicatable”, what are the steps in the stemming process?
(b) Explain cutoff method, entropy method, peak and plateau method. [4+12]
8. What are the problems with thesauri? Explain about statistical thesauri and theoretically thesauri. [16]

Code No: 07A71201

R07**Set No. 4**

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

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

Answer any FIVE Questions
All Questions carry equal marks

1. (a) Define the term handle. Explain a handle server.
 (b) What is RAP protocol? Where is it used. [10+6]
2. (a) What are the characteristics of clique approach?
 (b) Explain the single link clustering algorithm. What are the classes created when the algorithm is applied for creating clusters to Term relationship matrix? [4+12]
3. (a) Explain about measurements with automatic indexing.
 (b) How to define the measures with the search process? [8+8]
4. What were the reasons for origination of Information Retrieval Systems? What reasons forced to do research into Information Retrieval Systems? [16]
5. Explain the first use of n-grams. How n-grams is used in spelling error detection and correction? [16]
6. (a) Explain centroid comparisons with diagram.
 (b) Discuss the sources of potential errors in the final set of search terms from when a user first identifies a need for information to the creation of the final query. [6+10]
7. (a) What are the problems with Luhn's concept of "resolving power"?
 (b) Explain the techniques for creation of index when the terms of original item are used as basis of index process. [4+12]
8. (a) What are the sources of problems that arise in application of probability theory?
 (b) What are the trade offs in use of zoning as part of the indexing process? [4+12]

Code No: 07A71201

R07

Set No. 1

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

Time: 3 hours

Max Marks: 80

Answer any FIVE Questions
All Questions carry equal marks

1. Define the term Data Warehouse. What are the similarities between data warehouse and Information retrieval systems? [16]
2. What is the relationship between precision and TURR? [16]
3. (a) What is ranking? Explain about the relevance score.
(b) What is the purpose of collaborative filtering? [10+6]
4. (a) Where does the art of manual thesaurus construction reside?
(b) Explain about automatic term clustering. [4+12]
5. What are the approaches used to account for different document lengths while determining the value of term frequency? [16]
6. Which stemming technique is used by INQUERY system. Explain. [16]
7. Access the Internet and locate three information visualization techniques that are available. Describe what cognitive engineering principles are being used in the techniques. [16]
8. What are the goals of document summarization? [16]

Code No: 07A71201

R07

Set No. 3

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

Time: 3 hours

Max Marks: 80

Answer any FIVE Questions
All Questions carry equal marks

1. What is the use of Highlighting? Explain how highlighting is used in Boolean systems. [16]
2. Explain the importance of stemming algorithm. [16]
3. (a) What portions of an item should be indexed?
(b) Explain the weighting process of index terms.
(c) What is linkage? Explain precoordination and postcoordination. [4+6+6]
4. Explain the advantages of probabilistic approach. Give an example where the probabilistic approach may be applied. [16]
5. (a) What is a homograph?
(b) Explain about vocabulary constraints.
(c) How clustering effects precision and recall? [5+5+6]
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. Describe how other senses could be used in displaying results from searches. [16]
8. Assuming the database has the following words: act, able, arch, car, court, waste, wink, write, writer, wrinkle. Show the successor variety for the word "writeable". Apply the cutoff method, peak and plateau method and complete word method to determine possible stems for the word. Explain your rationale for cutoff method. [16]
