Code No: **R42059** **R10** **Set No. 1**

IV B.Tech II Semester Regular Examinations, April/May - 2014

INFORMATION RETRIEVAL SYSTEMS

(Computer Science and Engineering)

Time : 3 hours **Max. Marks: 75**

Answer any Five Questions

All Questions carry equal marks

*****

1 a) Define Information Retrieval System. Explain the objectives of the Information Retrieval System. [8]

b) Explain the functional overview of the Information Retrieval System. [7]

2 a) Apply the poster stemming algorithm the following words: irresponsible, informative, unrespectable [8]

b) Tradeoff the use of Precoordination versus Postcoordination [7]

3 a) What is the purpose of Thesaurus? Explain what it contains. [8]

b) Explain the concept of Inverted index file. How it can be used Information Retrieval. [7]

4 a) Write about PAT data structures and signature file structures [8]

b) What is automatic indexing? What are the various types of automatic indexing? [7]

5 a) Explain the concept of Information Extraction [8]

b) Differentiate Full Item indexing, Public File Indexing and Private File Indexing. [7]

6 a) Consider the following Term – Term matrix

	T1	T2	T3	T4	T5
T1	1	1	1	1	
T2	1	0	1	0	
T3	1	0	1	0	
T4	1	1	1	0	
T5	1	0	0	0	

Determine the clusters using Single Link and Star Techniques. [8]

b) What are the data files used to control and limit the stemming process by k-stem system. [7]

7 a) Explain Knuth – Prattt – Morris algorithm. [8]

b) Explain shift characters table. [7]

8 a) What algorithmetic basis is used for the GESCAN and Fast Data Finder hardware text search machines? [8]

b) Discuss about selective dissemination information search techniques with examples. [7]

Code No: **R42059** **R10** **Set No. 2**

IV B.Tech II Semester Regular Examinations, April/May - 2014

INFORMATION RETRIEVAL SYSTEMS

(Computer Science and Engineering)

Time : 3 hours **Max. Marks: 75**

Answer any Five Questions
All Questions carry equal marks
*****

1	a)	Explain the functional overview of the Information Retrieval System.	[8]
	b)	Clearly discuss the relevance of Information Retrieval Systems in the context of Digital libraries and Data Warehouses.	[7]
2	a)	Discuss the various Information Retrieval System capabilities in detail.	[8]
	b)	What the impact on precision and recall in the use is of stop list and stop algorithms?	[7]
3	a)	What is statistical indexing and what are the disadvantages of them?	[8]
	b)	Explain cutoff method, entropy method, peak and plateau method.	[7]
4	a)	Describe the similarities and differences between term stemming algorithms and n-grams.	[8]
	b)	Explain the following data structures giving suitable examples: i) N-gram ii) PAT	[7]
5	a)	Compare and contrast term clustering and item clustering.	[8]
	b)	Differentiate between Manual Clustering and Automatic Term Clustering. Explain with suitable examples.	[7]
6	a)	Explain statistical indexing in detail.	[8]
	b)	How do you expect that relevance feedback using negative judgments will affect the precision and recall of an information system.	[7]
7	a)	What are the data files used to control and limit the stemming process by k-stem system.	[8]
	b)	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?	[7]
8	a)	List out the differences between Boyer-Moore text search algorithm and Knuth-Pratt-Morris algorithm.	[8]
	b)	What algorithmetic basis is used for the GESCAN and Fast Data Finder hardware text search machines?	[7]

Code No: **R42059** **R10** **Set No. 3**

IV B.Tech II Semester Regular Examinations, April/May - 2014

INFORMATION RETRIEVAL SYSTEMS

(Computer Science and Engineering)

Time : 3 hours **Max. Marks: 75**

Answer any Five Questions
All Questions carry equal marks
*****

1	a)	Explain the functional overview of the Information Retrieval System.	[8]
	b)	What are the problems with Luhn's concept of "resolving power"?	[7]
2	a)	What is a Browse capability? Explain about various browse capabilities.	[8]
	b)	What is linkage? Explain precoordination and postcoordination.	[7]
3	a)	Which stemming technique is used by INQUERY system. Explain.	[8]
	b)	What are hypertext linkages?	[7]
4	a)	Explain the weighting process of index terms.	[8]
	b)	Describe the similarities and differences between term stemming algorithms and n-grams.	[7]
5	a)	Clearly bring out the steps of the process of clustering.	[8]
	b)	Define clustering. What are the general guidelines for clustering?	[7]
6	a)	Explain the advantages of probabilistic approach. Give an example where the probabilistic approach may be applied.	[8]
	b)	Write a short notes on the following i) Bayesian Model ii) Simple Term Frequency Algorithm	[7]
7	a)	Write short notes on the following with examples i) Similarity measures ii) Ranking algorithms	[8]
	b)	Discuss about selective dissemination information search techniques with examples.	[7]
8	a)	Describe the need for information visualization.	[8]
	b)	Differentiate hardware versus software text search algorithms.	[7]

Code No: **R42059** **R10** **Set No. 4**

IV B.Tech II Semester Regular Examinations, April/May - 2014

INFORMATION RETRIEVAL SYSTEMS

(Computer Science and Engineering)

Time : 3 hours **Max. Marks: 75**

Answer any Five Questions
All Questions carry equal marks
*****

1 a) Discuss the objectives of IRS [8]
b) What is a Browse capability? Explain about various browse capabilities. [7]

2 a) What is a Precoordination? How it is different from linkages? [8]
b) What are the problems with Luhn's concept of "resolving power"? [7]

3 a) Describe how the PAT Data structure is different from n-grams. [8]
b) What is indexing? Explain its objectives. Write about Indexing process. [7]

4 a) What is automatic indexing? Give the various classes of automatic indexing? [8]
b) Describe the similarities and differences between term stemming algorithms and n-grams. [7]

5 a) How clustering effects precision and recall? [8]
b) Compare and contrast manual clustering and Automatic Term Clustering. [7]

6 a) 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? [8]
b) Describe the need for information visualization. [7]

7 a) Consider the following Term – Term matrix [8]

	T1	T2	T3	T4	T5
T1	1		1	1	1
T2	1		0	1	0
T3	1		0	1	0
T4	1		1	1	0
T5	1		0	0	0

Determine the clusters using Cliques Techniques.
b) Discuss about selective dissemination information search techniques with examples. [7]

8 a) Differentiate hardware versus software text search algorithms. [8]
b) Write about Boyer-Moore text search algorithm. [7]