INSTITUTE OF AERONAUTICAL ENGINEERING

(Autonomous)
Dundigal, Hyderabad - 500 043

COMPUTER SCIENCE AND ENGINEERING

TUTORIAL QUESTION BANK

Course Name	INFORMATION RETRIEVAL SYSTEM
Course Code	A70533
Class	IV B. Tech I Semester
Branch	Computer Science and Engineering
Year	2018 – 19
Course Coordinator	Ms. S.J. Sowjanya Associate Professor, CSE
Course Faculty	Mr. N. V. Krishna Rao, Associate Professor, CSE Mr. C. Praveen Kumar, Assistant Professor, CSE

OBJECTIVES:

To meet the challenge of ensuring excellence in engineering education, the issue of quality needs to be addressed, debated and taken forward in a systematic manner. Accreditation is the principal means of quality assurance in higher education. The major emphasis of accreditation process is to measure the outcomes of the program that is being accredited.

In line with this, Faculty of Institute of Aeronautical Engineering, Hyderabad has taken a lead in incorporating philosophy of outcome-based education in the process of problem solving and career development. So, all students of the institute should understand the depth and approach of course to be taught through this question bank, which will enhance learner's learning process.

S No	Question	Blooms Taxonomy Level	Course Outcome
	UNIT – I		
	Part - A (Short Answer Questions)		
1	Define information retrieval system.	Remember	1
2	Give some differences between DBMS with information retrieval system.	Understand	1
3	How is browsing different from Searching?	Understand	1
4	Define your answer with relevant example Can information retrieval system be related to a database management system.	Remember	1
5	State briefly the terms 1. Precision 2. Recall	Remember	1
6	Differentiate the terms relevant and retrieved.	Understand	1
7	List three difference between data retrieval and information retrieval.	Remember	1
8	List five challenges of searching for information o the web.	Remember	1
9	Give some advantages of information retrieval system.	Understand	1
10	What is vector space model?	Remember	2
11	State Retrieval Strategy.	Remember	2
12	Define Smoothing.	Understand	2
13	How similarity coefficient can compute similarity between queries.	Understand	2

14	Describe the following statement in order to evaluate the effectiveness of a web search engine for general users, Would it generally be more important to measure precision or recall .	Remember	1
15	Differentiate digital libraries and data warehouses.	Remember	1
Part - B (Long Answer Questions)			
1	Explain the differences between Information Retrieval Systems and DBMS.	Understand	1
2	Give similarity coefficient and determine the ranking of following documents. Q:gold silver truck D1:shipment of gold damaged in a fire D2:delivery of silver arrived in a silver truck D3:shipment of gold arrived in a truck	Remember	2
3	Briefly explain the concept of simple term weights for the above query and documents.	Understand	2
4	Explain inverse document frequency.	Remember	1
5	List the objectives of IRS.	Understand	1
6	Discuss term frequencies with an example.	Remember	2
7	How is information retrieval system is related to Database management system.	Understand	2
8	Explain about the objectives of IRS.	Understand	2
9	State the concept of non binary independence model for the above query and documents.	Remember	1
10	Explain the concept smoothing for the above query and documents.	Understand	2
11	Discuss Similarities between vector space model and probabilistic retrieval strategy.	Understand	1
12	Explain the construction of vector document.	Remember	2
13	Calculate the similarity coefficient and determine the ranking of following documents in Probabilistic retrieval strategy. Q:gold silver truck D1:shipment of gold damaged in a fire D2:delivery of silver arrived in a silver truck D3:shipment of gold arrived in a truck	Understand	2
14	Discuss the term frequencies of the following. Q:new new times D1:new york times D2:new york posts D3:loss angels times	Understand	2
15	Define IDF and calculate the same for the above query and documents.	Remember	2
Part - C (Problem Solving and Critical Thinking Questions)			
1	Explain the Estimation of Document Vectors for the following 3 documents. D1:New York Times D2:New York Post D3: Los Angeles Times Q:New New Times	Understand	2
2	Give the use of invert index in vector space model.	Understand	1
3	Define Term weight.	Remember	1
4	Explain inverse document frequency.	Understand	2
5	Discribe about vector space model.	Remember	2
6	List different Retrieval Strategies.	Understand	2
7	Calculate the precision and recall scores for the search A Database contain 9 Records .A Search was Conducted on that Topic and 7 Records were retrieved .Of the 7 records retrieved,4 were relevant.	Understand	1

8	Calculate the precision and recall scores for the search A Database contain 80 Records .A Search was Conducted on that Topic and 60 Records were retrieved .Of the 60 records retrieved, 45 were relevant.	Understand	1
9	Explain the Estimation of Non-Binary independent model for the following three documents. D1:New York Times D2:New York Post D3: Los Angeles Times Q:New New Times	Remember	2
10	Give the Estimation of Smoothing in language model for the following three documents D1:New York Times D2:New York Post D3: Los Angeles Times Q:New New Times	Understand	2
UNIT – II			
Part - A (Short Answer Questions)			
1	Define the purpose of retrieval utilities.	Remember	3
2	State the concept of clustering as a retrieval utility.	Understand	3
3	Define how Relevance feedback is used to improve the results of retrieval strategy.	Remember	4
4	What is N-gram data structure?	Remember	5
5	Explain k-means algorithm.	Understand	5
6	Define term co-occurrence.	Remember	4
7	Remember six different sort orders to expand initial query in probabilistic model.	Remember	4
8	List three different bottom-up procedures used in hierarchically clustered collections.	Understand	3
9	Define regression analysis.	Remember	3
10	Discuss efficiency uses in clustering.	Understand	3
11	Give the formula for the basic weight in the probabilistic retrieval strategy.	Remember	4
12	List four variations for composing the new query.	Understand	4
13	Discuss three variations used in feedback iterations.	Remember	4
14	Write how users are involved in relevance feedback.	Understand	4
15	Define simple link clustering.	Understand	3
Part - B (Long Answer Questions)			
1	Explain about relevance feedback in vector space model.	Understand	3
2	Write about relevance feedback in probabilistic model.	Understand	3
3	Discuss the use of manually generated thesaurus.	Remember	5
4	Explain the concept of thesauri by constructing term-term similarity matrix.	Remember	3
5	Describe the approach of regression analysis to estimate the probability of relevance.	Understand	3
6	State how n-grams are used for detection and correction of spelling errors.	Remember	3
7	Define clustering and Explain hierarchical agglomerative clustering.	Understand	3
8	Explain the usage of document clustering to generate a thesaurus.	Remember	5
9	Give clustering with single value decomposition.	Remember	3
10	Write about term context used in thesaurus.	Remember	5
11	Discuss clustering without a recomputed matrix.	Understand	3
12	State extended relevance ranking with manual thesaurus.	Understand	4
13	Explain Rocchio and Buckshot clustering algorithm.	Understand	3
14	Give Damshek work for implementing five gram based measure of relevance.	Remember	4

15	Explain six different sort orders to expand initial query with the number of iterations to perform successful relevance feedback.	Understand	3
Part - C (Problem Solving and Critical Thinking Questions)			
1	Explain the use of probabilistic model in relevance feedback.	Remember	3
2	Write about n-gram developed by D Amore and Mah.	Understand	3
3	Describe clustering without pre computed matrix.	Remember	3
4	Differentiate single link clustering, complete linkage and group average.	Understand	5
5	Explain term co-occurrences in automatically constructed thesauri.	Understand	5
6	Describe about partial query expansion.	Remember	3
7	Explain vector space relevance feedback process.	Understand	3
8	Describe relevance feedback process with diagram.	Understand	4
9	Write about hierarchically clustered collections.	Remember	3
10	Discuss efficiency uses.	Understand	4
UNIT – III			
Part - A (Short Answer Questions)			
1	Discuss R-distance for calculating distance between query and document.	Understand	8
2	Describe how ranking is based on constrained spreading activation.	Remember	8
3	Remember how NLP is used to reduce ambiguity in language.	Remember	6
4	Describe cross language information retrieval.	Understand	7
5	What is query translation? Give an example.	Remember	6
6	Describe phrase translation.	Understand	7
7	State semantic network.	Remember	8
8	Define frame.	Remember	9
9	Describe slot.	Understand	6
10	Briefly explain spreading activation.	Remember	8
11	Explain the concept of pruning translation.	Understand	6
12	Define document translation.	Remember	8
13	Explain the approach of balancing queries.	Understand	7
14	Discuss about k-distance.	Remember	6
15	Describe evaluation of distance measures.	Remember	8
16	Give the performance of cross language information retrieval system.	Understand	6
17	Define parsing.	Understand	8
18	Discuss seven groups of relations into which a thesaurus is combined.	Understand	7
19	Explain the use of pivot language in translation.	Remember	9
20	Define stemming.	Understand	9
Part - B (Long Answer Questions)			
1	Explain the concept of semantic networks for automatic relevance ranking.	Understand	6
2	Why parsing is an essential feature of information retrieval system?	Understand	8
3	Describe three different types of translations.	Remember	7
4	Discuss unbalanced and structured queries approaches for choosing translations.	Understand	9
5	Explain about syntactic parsing.	Understand	8
6	Differentiate R-distance and K-distance.	Remember	7
7	Discuss balanced and pivot language approaches for choosing translations.	Remember	8
8	Explain what resources used to implement Cross language retrieval system.	Understand	8

9	Explain the measure to evaluate the performance of Cross language information retrieval system.	Understand	9
10	Discuss four questions to be answered to Cross language barrier.	Understand	9
11	Describe about four different approaches in choosing translations.	Remember	6
12	Explain how bilingual term list is used to improve accuracy.	Understand	7
13	State the use of POS word sense tagging.	Understand	9
14	Describe how message understanding conference focuses on information extraction.	Remember	8
15	Explain the concept of distance measures in a semantic network.	Understand	7
Part - C (Problem Solving and Critical Thinking Questions)			
1	Discuss R-distance and K-distance.	Understand	7
2	Explain simple phrases and complex phrases.	Understand	8
3	Describe balanced query and structured query.	Remember	8
4	Discuss about unbalanced queries.	Understand	8
5	Write about quality of bilingual term lists.	Understand	7
6	Describe the method used to translate a query.	Understand	9
7	Explain the measures used to evaluate the performance of cross-language information retrieval systems.	Understand	8
8	Describe the resources used to implement cross-language information retrieval systems.	Remember	9
9	Discuss ranking based on constrained spreading activation.	Understand	8
10	Describe developing query term based on concepts.	Understand	9
UNIT - IV			
Part - A (Short Answer Questions)			
1	Define index pruning.	Remember	10
2	Remember posting list.	Remember	12
3	Define document file.	Understand	12
4	What is index?	Understand	11
5	Write about I-Match.	Remember	10
6	Give the method to find exact duplicates.	Understand	11
7	State how scanning removes false positives.	Understand	12
8	List two advantages of index file.	Remember	10
9	Classify different types of files.	Remember	10
10	Define weight file.	Understand	11
11	Remember about two top-down algorithms.	Remember	12
12	List index compression algorithms.	Understand	11
13	Define Fixed length Index Compression.	Remember	12
14	State variable length index compression.	Understand	10
15	Remember about cutoff based on document frequency.	Remember	12
Part - B (Long Answer Questions)			
1	Explain methods to reorder documents prior to indexing.	Understand	10
2	Discuss methods to compress an inverted index.	Remember	10
3	Define efficiency. Explain about inverted index.	Remember	11
4	Write about throughput-optimized compression.	Understand	12
5	List various top-down and bottom-up algorithms.	Understand	12
6	Define how inverted index allows quick search of a posting list.	Remember	11
7	Explain about duplicate document detection.	Understand	10
8	Describe method to build an inverted index.	Understand	12
9	Give the method for finding similar duplicates.	Remember	12
10	Explain how signature files are used to detect duplicates.	Understand	12
11	List three methods to characterize posting list.	Understand	10
12	Define about query processing.	Remember	12
13	Discuss about partial result set retrieval.	Understand	12

14	Explain about I-match used in duplicate document detection.	Understand	12
15	Define vector space simplifications.	Remember	11
Part - C (Problem Solving and Critical Thinking Questions)			
1	Explain about Digital Libraries and Data Warehouses.	Understand	12
2	Differentiate "Digital Library" and an Information Retrieval System. What new areas of information retrieval research may be important to support a Digital Library.	Understand	12
3	List different Browsing Capabilities.	Remember	12
4	Define Indexing. Explain the objectives of indexing and also discuss about Automatic indexing.	Understand	10
5	List two major data structures in any information system.	Understand	11
6	Describe the similarities and differences between term stemming algorithms and n-grams.	Remember	11
7	Explain in detail about Vector Weighting. What are the general problems with the Vector Model.	Remember	12
8	Write about Natural Language Processing. Describe how use of Natural Language Processing will assist in the disambiguation process.	Understand	11
9	Explain Similarity Measures and Ranking.	Understand	12
10	Describe two major approaches to generating queries. Explain in Detail .	Remember	12
UNIT - V			
Part - A (Short Answer Questions)			
1	Define Data Integrity.	Remember	13
2	What is performance?	Understand	14
3	State Portability.	Understand	15
4	Remember the extensions to SQL.	Remember	14
5	List different types of User-defined Operators.	Understand	15
6	List NFN Approaches.	Remember	14
7	Define proximity searches works.	Remember	15
8	List the operators used in Boolean query.	Understand	14
9	State Boolean Retrieval.	Remember	13
10	Define Relational Information Retrieval system.	Understand	15
11	Write about Relational Schema.	Remember	14
12	Give a method to store XML Metadata.	Understand	13
13	Discuss about XML-QL.	Remember	15
14	What is an Index?	Understand	14
15	Define attributes in Index.	Remember	14
Part - B (Long Answer Questions)			
1	Explain about historical progression.	Understand	15
2	Discuss briefly about user-defined operators.	Understand	13
3	List Non-first normal form approaches.	Remember	15
4	Discuss about information retrieval as a relational application.	Understand	13
5	Explain about Boolean queries.	Understand	15
6	Describe about proximity searches.	Remember	14
7	Write about the computation of relevance using unchanged SQL.	Understand	13
8	Describe semi-structured search using a relational schema.	Understand	15
9	State how static relational schema support XML-QL.	Remember	14
10	Discuss about relational information retrieval system.	Understand	14
11	Explain the method of tracking XML documents.	Remember	13
12	Give the method that how index table models an XML index.	Understand	15
13	Explain about a theoretical model of distributed retrieval.	Understand	14
14	Describe centralized information retrieval system model.	Understand	15
15	Write about distributed information retrieval system model.	Remember	13
Part - C (Problem Solving and Critical Thinking Questions)			
1	Discuss evaluation of web search engines.	Remember	13

2	Describe how run time performance is a disadvantage of information retrieval.	Understand	15
3	Explain how information retrieval becomes relational application.	Remember	14
4	Write about relevance ranking?	Understand	15
5	State how XML has become the standard for platform– independent data exchange.	Remember	14
6	Explain how data integrity and portability are disadvantages of information retrieval.	Understand	15
7	State how semi structured search is performed using relational schema.	Remember	14
8	List two methods of distributed retrieval.	Remember	15
9	Describe briefly about web search.	Understand	14
10	Give the method to improve effectiveness of web search engines.	Remember	13

Prepared By**:** Ms. S.J. Sowjanya , Associate Professor, CSE.
Mr. N.V. Krishna Rao, Associate Professor, CSE.
Mr. C. Praveen Kumar, Assistant Professor, CSE.

HOD, COMPUTER SCIENCE AND ENGINEERING