SCHEME OF COURSE WORK

Course Details:

Course Title	INFORMATION RETREIVAL SYSTEMS		
Course Code	15IT2115	L T P C	:4 0 0 3
Program:	M.Tech.		
Specialization:	SOFTWARE ENGINEERING		
Semester	I		
Prerequisites	None		
Courses to which it is a prerequisite	None		

Course Outcomes (COs):

1	Identify pre-processing methods for effective information retrieval
2	Apply tolerant information retrieval
3	Describe the index compression process
4	Transform textual information into vectors
5	Analyze ranked and unranked search results

Program Outcomes (POs):

A graduate of Information Technology will be able to

1	Demonstrate in-depth knowledge of Software Engineering with analytical and synthesizing skills.
2	Analyze complex problems critically and provide viable solution.
3	Evaluate potential solutions to a problem and arrive at optimal solutions
4	Apply research methodologies to develop innovative techniques for solving complex Information Technology related problems.
5	Apply techniques and tools to solve complex problems.
6	Effective team member in a collaborative and multidisciplinary project to achieve common goal
7	Manage a software team and to maintain financial records as per standard
8	Effectively communicate with clients, peers and society at large
9	Take up lifelong learning to be in tune with the fast-changing software related technologies
10	Follow ethical practices in the software industry and accept social responsibility
11	Learn independently from mistakes and surge forward with positive attitude and enthusiasm.

Course Outcome Versus **Program Outcomes:**

COs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO-1	M	M										
CO-2		M			S				M			
CO-3												
CO-4				M								
CO-5				M								

S - Strongly correlated, *M* - *Moderately* correlated, *Blank* - *No correlation*

Assessment Methods:	Seminar / Mid-Test / End Exam

Teaching-Learning and Evaluation

Week	TOPIC / CONTENTS	Course Outcomes	Sample questions	TEACHING-LEARNING STRATEGY	Assessment Method & Schedule
1	**UNIT-I: Boolean Retrieval:** An example information retrieval problem, A first take at building an inverted index,	CO1	Define the term Boolean Query.	▫ Lecture	Mid –Test 1
2	Processing Boolean queries, The extended Boolean model versus ranked retrieval.	CO1	What are the advantages of extended Boolean model over traditional model?	▫ Lecture	Mid –Test 1
3	**The Term vocabulary and postings lists :** Document delineation and character sequence decoding, Obtaining the character sequence in a document, Choosing a document unit, Determining the vocabulary of terms.	CO1	How to define a document unit?	▫ Lecture	Mid –Test 1
4	Tokenization, Dropping common terms: stop words, Normalization (equivalence classing of terms)	CO1	Explain Token Normalization	▫ Lecture / ▫ Demonstration	Mid –Test 1
5	Stemming and lemmatization, Faster postings list intersection via skip pointers,	CO1	How stemming is helpful in information retrieval?	▫ Lecture / ▫ Demonstration	Mid –Test 1
6	Positional postings and phrase queries , Biword indexes , Positional indexes , Combination schemes	CO1	What is meant by Biword indexes?	▫ Lecture	Mid –Test 1 Seminar
7	**UNIT II: Dictionaries and tolerant retrieval :** Search structures for dictionaries, Wildcard queries, General wildcard queries, k-gram indexes for wildcard queries, Spelling correction.	CO2	What are the different search statements for dictionaries	▫ Lecture	Mid –Test 1 Seminar
8	Implementing spelling correction, Forms of spelling correction , Edit distance , k-gram indexes for spelling correction, Context sensitive spelling correction , Phonetic correction.	CO2	What is isolated-term correction and context-sensitive correction?	▫ Lecture	Mid –Test 1 Seminar
9	**Mid-Test 1**	CO1 & CO2			Mid-Test 1 (Week 9)
10	**Index construction :** Hardware basics , Blocked sort-based indexing, Single-pass in-memory indexing , Distributed indexing , Dynamic indexing , Other types of indexes	CO2	What are the hardware issues associated with indexing?	▫ Lecture / ▫ Demonstration	Mid –Test 2
11	**UNIT –III: Index compression:** Statistical properties of terms in information retrieval, Heaps' law: Estimating the number of terms , Zipf's law:Modeling the distribution of terms ,	CO3	What is Zipf's law?	▫ Lecture ▫ Problem solving	Mid –Test 2 Seminar
12	Dictionary compression, Dictionary as a string , Blocked storage , Postings file compression, Variable byte codes , ã codes	CO3	What is the significance of variable bye encoding?	▫ Lecture	Mid –Test 2 Seminar
13	**Scoring, term weighting :** Parametric and zone indexes, Weighted zone scoring, Learning weights, The optimal weight g, Term frequency and weighting , Inverse document frequency, Tf-idf weighting.	CO3	Write briefly about inverse document frequency	▫ Lecture / ▫ Demonstration	Mid –Test 2 Seminar
14	**UNIT –IV: The vector space model:** The vector space model for scoring, Dot products , Queries as vectors , Computing vector scores, Variant tf-idf functions ,	CO4	What are the weighting functions/schemes used in variant idf functions	= Lecture	Mid –Test 2
15	Sublinear tf scaling, Maximum tf normalization, Document and query weighting schemes, Pivoted normalized document length.	CO4	List some of the query weighting schemes.	= Lecture	Mid –Test 2
16	**UNIT –V : Evaluation in information retrieval :** Information retrieval system evaluation, Standard test collections , Evaluation of unranked retrieval sets , Evaluation of ranked retrieval results,	CO5	Discuss about Evaluation of unranked retrieval sets.	▫ Lecture	Mid-Test 2
17	Assessing relevance , Critiques and justifications of the concept of Relevance, A broader perspective: System quality and user utility, System issues, User utility, Refining a deployed system, Results snippets	CO5	How do you measure the relevance of retrieved results? What is a snippet?	▫ Lecture / ▫ Demonstration	Mid-Test 2
18	Mid-Test 2	CO3, CO4, CO5			Mid-Test 2
19/20	END EXAM				