

Subject Code : 2CS2010209	Subject Title: INFORMATION RETRIEVAL (IR)
Pre-requisite :	Basic knowledge of web designing, Data Structure.

Course Objective:

The objectives of the course are to:

- Understand concepts of Information Retrieval.
- Understand functions of an information retrieval system and analyze the components of an information retrieval system.
- Understand concept of textual document indexing, indexing and searching, Human Computer Interaction, text analytics and retrieval evaluations.
- Understand the concept of Web Retrieval and Digital Libraries.

Teaching Scheme (Hours per week)				Evaluation Scheme (Marks)				
Lecture	Tutorial	Practical	Credit	Theory		Practical		Total
				University Assessment	Continuous Assessment	University Assessment	Continuous Assessment	
3	-	3	6	60	40	30	20	150

Subject Contents				
Sr. No	Topic	Total Hours	Weight (%)	
1	Introduction, Modeling Motivation, Basic Concepts, Retrieval Process, Taxonomy of Information Retrieval Models, Classic Information Retrieval, Structured Text Retrieval Model	15	6	
2	Retrieval Evaluation, Query Languages, Query Operations Introduction, Performance Evaluation, Reference Collections, Keyword Based Querying, Pattern Matching, Structural Queries, Query Protocols.	25	10	
3	Text and Multimedia Language and Properties, Text Operations, Indexing and Searching Meta-data, Text, Markup Languages, Multimedia, Document Preprocessing, Document Clustering, Text Compression, Comparing Text Compression Techniques, Inverted Files, Other Indices for Text, Boolean Queries	25	10	
4	User Interface and Visualization, Searching the Web Human Computer Interaction, Information Access Process, Query Specification, Challenges, Characterizing the Web, Search Engines, Browsing, Meta-searches, Searching using Hyper-links	20	8	
5	Libraries and Bibliographical Systems, Digital Libraries Online IR Systems and Document Databases, Online Public Access Catalogs (OPACs), Digital Libraries: Architectural Issues, Document Models, Representations and Access.	15	8	

Course Outcome:

At the end of this course, the student would be able

- To learn fundamental concept of Information Retrieval (IR), basic algorithms and techniques for information retrieval (document indexing and retrieval, query processing).
- To learn various types of query languages, text operations and apply quantitative evaluation methods for the IR systems

- To explain different retrieval models and basic algorithms involved in processing and retrieval of information.
- To evaluate existing information retrieval systems and suggest how the systems can be Improved.

List of References:

1. Ricardo Baeza-Yates, Berthier Ribeiro-Neto "Modern Information Retrieval", Pearson India Education, Inc., 2007. Copyright by the ACM Press. 513 pp., ISBN: 978-81-317-0977-1
2. C. Manning, P. Raghavan, and H. Schütze, "Introduction to Information Retrieval", Cambridge University Press, 2008
3. Pang-Ning Tan, Michael Steinbach and Vipin Kumar, "Introduction to Data Mining", Addison-Wesley, 2005
4. Robert R. Korfhage, "Information Storage and Retrieval", John Wiley & Sons, 1997, 349 pp., ISBN: 0-471-14338-3

List of Experiments:

Note: The experiment list provided beneath is for reference only. The course teacher may Change/formulate it as per his/her methodology and requirement.

Sr.No	Practical Experiments
1.	A Search Engine for all pages of your college domain, crawl index and search all pages in your college domain.
2.	Link Mining in your college Domain, Analyze the hyperlink structure of the College domain.
3.	Metsearch Engine, Build a search engine which combines the result of 4 other popular search engines.
4.	Hypertext Clustering, Cluster hypertext documents.
5.	Implementation of various classification algorithms on text.