

Branch Name:	Information and Communication Technology
Program Code:	ET107
Semester/Year:	Semester VI/ Third Year
Course Name:	Information Retrieval (Professional Elective-III)
Course Code:	2ET1070603T
Pre-requisite:	Basics of Machine Learning

Course Objective:

1. Learn to write code for text indexing and retrieval.
2. Learn to evaluate information retrieval systems
3. Learn to analyze textual and semi-structured data sets
4. Learn to evaluate information retrieval systems
5. Learn about text similarity measure
6. Understanding about search engine
7. Text Classification

Teaching and Examination Scheme:

Credit	Teaching Scheme Hours per week				Evaluation Scheme (Marks)								Total (Marks)
					Theory (Marks)				Practical (Marks)				
	L	T	P	Total	SEE		CCE		SEE		CCE		
					Min.	Max.	Min.	Max.	Min.	Max.	Min.	Max.	
3	3	-	-	3	24	60	16	40	-	-	-	-	100

Subject Contents:

Sr. No.	Topic	Total Hrs.	Weightage %
1	Overview of text retrieval systems Information Retrieval, Early Developments, The IR Problem, The Users Task, Information versus Data Retrieval, The IR System, The Software Architecture of the IR System, The Retrieval and Ranking Processes, The Web, The e-Publishing Era, How the web changed Search, Practical Issues on the Web, How People Search, Search Interfaces Today, Visualization in Search Interfaces.	9	17
2	Retrieval models and implementation: Vector Space Models Vector Space Model, TF-IDF Weight, Evaluation in information retrieval	2	8
3	Query expansion and feedback Relevance feedback, pseudo relevance feedback, Query Reformulation	5	14
4	Probabilistic models; statistical language models Okapi/BM25; Language model, KL-divergence, Smoothing	8	15
5	Text classification & Text clustering The text classification problem, Naive Bayes text classification, k- nearest neighbors, Support vector Machine, Feature Selection, Vector-space	10	20

	clustering; K-means algorithm, Hierarchical clustering, DBSCAN algorithm PAM and PAMK, EM algorithm		
6	Web Retrieval and Web Crawling The Web, Search Engine Architectures, Cluster based Architecture, Distributed Architectures, Search Engine Ranking, Link based Ranking, Simple Ranking Functions, Learning to Rank, Evaluations, Search Engine Ranking, Search Engine User Interaction, Browsing, Applications of a Web Crawler, Taxonomy, Architecture and Implementation, Scheduling Algorithms, Evaluation.	8	15
7	IR applications Information extraction, Question answering, Opinion summarization, Social Network	3	11

List of Reference Books:

1. Ricardo Baeza-Yates and Berthier Ribeiro-Neto, —Modern Information Retrieval: The Concepts and Technology behind Search, Second Edition, ACM Press Books, 2011.
2. Ricci, F, Rokach, L. Shapira, B.Kantor, —Recommender Systems Handbook, First Edition, 2011.
3. C. Manning, P. Raghavan, and H. Schütze, —Introduction to Information Retrieval, Cambridge University Press, 2008.
4. Stefan Buettcher, Charles L. A. Clarke and Gordon V. Cormack, —Information Retrieval: Implementing and Evaluating Search Engines, The MIT Press, 2010.

Course Learning Outcomes (CLO): On completion of this course, the students will be able to:

CLO	Description	Bloom's Taxonomy Level
CLO1	Understand comprehend the key concepts of clustering algorithm	1. Remembering 2. Understanding
CLO2	Learn representation of Various track at CLEF 2015	2 Understanding 5. Evaluate
CLO3	Learn IR from Micro blog	4. Analyze
CLO4	Learn Implement Domain specific Search Engine	2 Understanding 5. Evaluate

Mapping of CLOs with POs & PSOs

Course Learning Outcomes	Program Outcomes (POs)												Program Specific Outcomes(PSOs)			
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO 10	PO 11	PO 12	PSO 1	PSO 2	PSO 3	PSO 4
CLO1					L	H	L				L	H	H	M	L	
CLO2	H		M	M	L					L	L	M		M	L	
CLO3		M	L		L		L							H		L
CLO4	H	M			L			L					L		M	H

H: High, M: Medium, L: Low