Course Title	Information Retrieval Systems	Course No *(will be assigned)*			
Specialization	Computer Engineering	Structure (IPC)	3	0	3
Offered for	UG & DD	Status	Core ☐	Elective ☐	
Faculty	B Siva Selvan	Type	New ☐	Modification ☐	
Pre-requisite	COM 207T, COM 212T	To take effect from	Jan 2016		
Submission date	November 2016	Date of approval by Senate			
Objectives	The course explores the underlying theories and algorithms of information retrieval systems and introduces the methodology for the design and evaluation of information retrieval systems.				
Course outcomes	Students will be able to understand underlying concepts such as data structures for unstructured data, indexing, retrieval, etc. and shall be equipped with skills required to design, develop and use information retrieval and search systems.				
Contents of the course	Text operations before indexing, such as stop word removal, stemming Crawling, collecting web documents. Surface web crawling and deep web crawling. Crawling tools. Near duplicate detection. MinHash algorithm Indexing, constructing an inverted index of word to document pointers Searching, retrieving documents that contain a given query token from the inverted index Use of Lucene to construct a practical and large search engine Various retrieval models, including the classic boolean model and vector space model TF-IDF and their variants - Statistic properties of text, power laws, Zipf's law, Heaps' law Language models, unigram model and bigram model. Evaluation criteria, precision and recall, F1 Document clustering and classification. Naive Bayes and support vector machine. Feature selection. Mutual information - Ranking, scoring retrieved documents according to relevance metrics. PageRank and HITs algorithms.				
Textbook	1. Christopher D. Manning, Prabhakar Raghavan and Hinrich Schutze,Introduction to Information Retrieval, Cambridge University Press. 2008.				
References	1. W Bruce Croft, Donald Metzler, and Trevor Strohman. Search engines: Information retrieval in prac- tice. Pearson, 2009. 2. R B Yates, Neto, Modern Information Retrieval: The Concepts and Technology behind Search, Addison Wesley, 2011. 3. S Chakrabarti, Mining the Web: Discovering Knowledge from Hypertext Data – Morgan Kauffman, 2002				