# RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA, BHOPAL

## New Scheme Based On AICTE Flexible Curricula

## Artificial Intelligence and Data Science, VI-Semester

## Departmental Elective – AD 603 (C) Information Retrieval

Course Outcomes: After completion of the course students will be able to

CO1: Understand the role of Information Retrieval on Web and Artificial Intelligence.
CO2: Comprehend the different space retrieval models in IR and pre-processing.
CO3: Able to understand the web architectures and crawling techniques.
CO4: Able to understand Link Analysis for Information Retrieval.
CO5: Able to elaborate text mining, Classification and clustering algorithms

Unit-I

Introduction - History of IR- Components of IR - Issues -Open source Search engine. Frameworks - The Impact of the web on IR - The role of artificial intelligence (AI) in IR – IR Versus Web Search - Components of a search engine, characterizing the web.

Unit -II

Boolean and Vector space retrieval models- Term weighting - TF-IDF weighting- cosinesimilarity - Pre-processing - Inverted indices - efficient processing with sparse vectors LanguageModel based IR - Probabilistic IR -Latent Semantic indexing - Relevance feedback and queryexpansion.

Unit- III

Web search overview, web structure the user paid placement search engine optimization, Web
Search Architectures - crawling - meta-crawlers, Focused Crawling - web indexes - Near duplicate detection - Index Compression - XML retrieval.

Unit -IV

Link Analysis -hubs and authorities - Page Rank and HITS algorithms -Searching and Ranking
Relevance Scoring and ranking for Web - Similarity - Hadoop & Map Reduce - Evaluation - Personalized search - Collaborative filtering and content-based recommendation of documents
And products - handling invisible Web - Snippet generation, Summarization. QuestionAnswering, Cross-Lingual Retrieval.

Unit -V

Information filtering: organization and relevance feedback - Text Mining- Text classification andclustering - Categorization algorithms, naive Bayes, decision trees and nearest neighbor - Clustering algorithms: agglomerative clustering, k-means, expectation maximization (EM).

References:
1. C. Manning, P. Raghvan and H Schutze: Introduction to Information Retrieval, Cambridge University Press, 2008.
2. Ricardo Baeza -Yates and Berthier Ribeiro –Neto, Modern Information Retrieval The Concepts and Technology behind Search 2nd Edition, ACM Press Books 2011.
3. Bruce Croft, Donald Metzler and Trevor Strohman Search Engines Information Retrieval in Practice 1st Edition Addison Wesley, 2009
4. MarkLevene, An Introduction to Search Engines and Web Navigation, 2nd Edition Wiley 2010.