

Course Code	Course Name	Hours per Week			Total	Total
		L	T	P	Hrs.	Credits
CS5CD02	Information Retrieval	2	0	0	2	2

UNIT-I

Taxonomy of IR Models, Classic models, Set theoretic model, Algebraic models, Probabilistic model, Structured text retrieval models, Models for browsing, Retrieval evaluations, Reference collections

UNIT-II

Query Languages, Query Operations, Text and Multimedia Languages, Text Operations, Document Preprocessing, Matrix Decompositions and Latent Semantic Indexing, Text Compression, Indexing and Searching, Inverted Files, Suffix Trees, Boolean Queries, Sequential Searching, Pattern Matching

UNIT-III

Text Classification, Naïve Bayes Vector Space Classification, Support Vector Machines and Machine Learning on Documents, Flat Clustering, Hierarchical Clustering.

UNIT- IV

Web Search Basics, Web Characteristics, Index size and Estimation, Near duplicates and Shingling, Web Crawling, Distributing Indexes, Connectivity Servers, Link Analysis, Web as a Graph, Page Rank, Hubs and Authorities, Question Answering.

UNIT-V

Online IR Systems, Online Public Access Catalogs, Digital Libraries, Architectural Issues, Document Models, Representations and Access, Protocols

TEXT BOOKS

1. Modern Information Retrieval: The Concepts and Technology behind Search (2nd Edition): Ricardo Baeza Yates, Berthier RibeiroNeto, Addison Wesley.
2. Introduction to Information Retrieval (1st Ed): Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, Cambridge University Press.
3. Search Engines: Information Retrieval in Practice (1st Ed): Bruce Croft, Donald Metzler and Trevor Strohman, , Addison Wesley.

REFERNCES

1. Introduction to Information Retrieval: Reghavan, Cambridge University Press; South Asian Edition.
2. Modern Information Retrieval: Recardo Baeza-Yates, Pearson Education.