

CP5094 INFORMATION RETRIEVAL TECHNIQUES

DETAILED SYLLABUS

OBJECTIVES

- To understand the basics of information retrieval with pertinence to modeling, query operations and indexing
- To get an understanding of machine learning techniques for text classification and clustering.
- To understand the various applications of information retrieval giving emphasis to multimedia IR, web search
- To understand the concepts of digital libraries

UNIT I INTRODUCTION: MOTIVATION

Basic Concepts – Practical Issues - Retrieval Process – Architecture - Boolean Retrieval – Retrieval Evaluation – Open Source IR Systems–History of Web Search – Web Characteristics– The impact of the web on IR —IR Versus Web Search–Components of a Search engine

UNIT II MODELING

Taxonomy and Characterization of IR Models – Boolean Model – Vector Model - Term Weighting – Scoring and Ranking –Language Models – Set Theoretic Models - Probabilistic Models – Algebraic Models – Structured Text Retrieval Models – Models for Browsing

UNIT III INDEXING

Static and Dynamic Inverted Indices – Index Construction and Index Compression. Searching - Sequential Searching and Pattern Matching. Query Operations -Query Languages – Query Processing - Relevance Feedback and Query Expansion - Automatic Local and Global Analysis – Measuring Effectiveness and Efficiency

UNIT IV CLASSIFICATION AND CLUSTERING

Text Classification and Naïve Bayes – Vector Space Classification – Support vector machines and Machine learning on documents. Flat Clustering – Hierarchical Clustering –Matrix decompositions and latent semantic indexing – Fusion and Meta learning

UNIT V SEARCHING THE WEB

Searching the Web –Structure of the Web –IR and web search – Static and Dynamic Ranking
– Web Crawling and Indexing – Link Analysis - XML Retrieval Multimedia IR: Models and
Languages – Indexing and Searching Parallel and Distributed IR – Digital Libraries

REFERENCES

1. Christopher D. Manning, Prabhakar Raghavan, Hinrich Schutze, —Introduction to Information Retrieval, Cambridge University Press, First South Asian Edition, 2008.
2. Implementing and Evaluating Search Engines, The MIT Press, Cambridge, Massachusetts London, England, 2010
3. Ricardo Baeza – Yates, Berthier Ribeiro – Neto, —Modern Information Retrieval: The concepts and Technology behind Search (ACM Press Books), Second Edition, 2011.
4. Stefan Butcher, Charles L. A. Clarke, Gordon V. Cormack, —Information Retrieval