CS8080 INFORMATION RETRIEVAL TECHNIQUES

DETAILED SYLLABUS

OBJECTIVES:

- To understand the basics of Information Retrieval.
- To understand machine learning techniques for text classification and clustering.
- To understand various search engine system operations.
- To learn different techniques of recommender system.

UNIT I INTRODUCTION

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.

UNIT II MODELING AND RETRIEVAL EVALUATION

Basic IR Models - Boolean Model - TF-IDF (Term Frequency/Inverse Document Frequency) Weighting - Vector Model – Probabilistic Model – Latent Semantic Indexing Model – Neural Network Model – Retrieval Evaluation – Retrieval Metrics – Precision and Recall – Reference Collection – User-based Evaluation – Relevance Feedback and Query Expansion – Explicit Relevance Feedback.

UNIT III TEXT CLASSIFICATION AND CLUSTERING

A Characterization of Text Classification – Unsupervised Algorithms: Clustering – Naïve Text Classification – Supervised Algorithms – Decision Tree – k-NN Classifier – SVM Classifier – Feature Selection or Dimensionality Reduction – Evaluation metrics – Accuracy and Error – Organizing the classes – Indexing and Searching – Inverted Indexes – Sequential Searching – Multi-dimensional Indexing.

UNIT IV 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.

UNIT V RECOMMENDER SYSTEM

Recommender Systems Functions – Data and Knowledge Sources – Recommendation Techniques – Basics of Content-based Recommender Systems – High Level Architecture – Advantages and Drawbacks of Content-based Filtering – Collaborative Filtering – Matrix factorization models – Neighbourhood models.

TEXT 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.

REFERENCES:

1. C. Manning, P. Raghavan, and H. Schütze, —Introduction to Information Retrieval, Cambridge University Press, 2008.

2. Stefan Buettcher, Charles L. A. Clarke and Gordon V. Cormack, —Information Retrieval: Implementing and Evaluating Search Engines, The MIT Press, 2010.