

PREREQUISITES:

- An introduction to database systems, covering SQL and related programming
- A sophomore-level course in Data Structures, Algorithms, and Discrete Mathematics

OBJECTIVES:

- Study the Web as an object
- Improving Web applications
- Social Mining

SYLLABUS:

Unit I: Web Crawling - Algorithms - Implementation Issues - Types - Crawler Ethics and Conflicts - Indexing - Text Analysis and Classification

Unit II: Ranking Algorithms - Web Search and Retrieval - Link Analysis - Social Network Analysis - Co-Citation and Bibliographic Coupling - Page Rank, HITS - Community Discovery - Term Frequency - Inverse Document Frequency (TF.IDF)

Unit III: Building DOM Trees - Extraction Based on Multiple Pages - Information Integration - Schema Matching - Pre-Processing for Schema Matching - Strategies for Combining Similarities - Integration of Web Query interfaces - Constructing a Unified Global Query Interface

Unit IV: Clustering/Community Algorithms - Topical Locality - Web growth models - Web Traffic models

Unit V: Social Tagging - Social Networks and Social Media - Information Diffusion

REFERENCES:

Anand Rajaraman, Jure Leskovec and Jeffrey D. Ullman. **Mining of Massive Data Sets.** Cambridge University Press, 2012.

Jiawei Han and Micheline Kamber. **Data Mining: Concepts and Techniques.** Morgan Kaufmann Publishers, 2006.

Soumen Chakrabarti. **Mining the Web: Discovering Knowledge from Hypertext Data.** Morgan Kaufmann Publishers, 2003.

Bing Liu. **Web Data Mining, Exploring Hyperlinks, Contents and Usage Data.** Springer, 2007.

Shweta
referred
approved
M/C

41

Mrs - RB
M/C
T.C.