



STANLEY

COLLEGE OF ENGINEERING & TECHNOLOGY FOR WOMEN
(Approved by AICTE , New Delhi | Affiliated to Osmania University ,Hyderabad)
Address : Chapel Road, Abids ,Hyderabad

CS578

WEB MINING

UNIT-I

Crawling and Indexing, Topic Directories, Clustering and Classification. Hyperlink analysis.Resource Discovery and Vertical Portals, Structured vs Unstructured Data Mining.

Crawling the web: HTML and HTTP basics, Crawling Basics, Engineering Large Scale Crawlers, Putting Together a crawler

UNIT-II

Web Search and Information Retrieval: Boolean Queries and Inverted index, Relevance Ranking, Similarity Search

Similarity and Clustering: Foundations and Approaches, Bottom-up and Top-Down partition paradigms.

UNIT-III

Supervised learning: Introduction, Overview, of classification strategies, Nearest Neigh Learners, Feature Selection, Bayesian Learners, Discriminative classification, Hypertext classification

UNIT-IV

Expectation Maximization, Labelling Hypertext Graphs, Co-Training Social Sciences and bibliometry, Page Rank and HITS, Coarse Grained Graph Model, Enhanced Model and Techniques, Evaluation of Topic Distillation.

UNIT-V

Collecting Important Pages, Similarity Search using Link Topology, Topical Locality and Focused Crawling, Discovering Communities.

Future of Web Mining: Information Extraction, Natural Language Processing, Question Answering, Profile, Personalization, and Collaboration.

Suggested Reading:

1. Chakrabarti Soumen, 'Mining the Web: Discovering Knowledge From Hypertext Data", Boston Elsevier, 2003.
2. Manu Konchady, "Text Mining Application Programming" Cengage Learning, 2006.