SREENIVASA INSTITUTE of TECHNOLOGY and MANAGEMENT STUDIES

III MCA - I Semester	**L**	**P**	**C**
	4	**0**	**4**

16MCA314B **WEB MINING**

Course Objectives:

- To focus on a detailed overview of the data mining process and techniques, specifically those that are relevant to Web mining.
- To Understand the basics of Information retrieval and Web search with special emphasis on web Crawling.
- To appreciate the use of machine learning approaches for Web Content Mining.
- To understand the role of hyper links in web structure mining.
- To appreciate the various aspects of web usage mining.

Syllabus:

UNIT I : Introduction

Introduction: What is the World Wide Web? . A Brief History of the Web and the Internet . What is Web Mining?.

Information retrieval and Web search: Information retrieval Models -- Relevance Feedback - Text and Web page Pre-processing – Inverted Index – Latent Semantic Indexing – Web Search – Meta-Search – Web spamming

UNIT II : Web Content Mining

Supervised Learning: Decision tree - Naïve Bayesian Text Classification - Support Vector Machines K-Nearest Neighbourhood Learning - Classifier Evaluation

Unsupervised Learning: K-means Clustering - Hierarchical Clustering – Distance Functions – Cluster Evaluation.

UNIT III : Web Link Mining

Link Analysis: Social Networks Analysis – Co-Citation and Bibliographic Coupling - Page Rank – HITS .

Web Crawling: A Basic Crawler Algorithm - Implementation Issues - Universal Crawlers - Focused Crawlers - Topical Crawlers - Evaluation

UNIT IV : Structured Data Extraction

Wrapper Generation

Structured Data Extraction:Wrapper Generation: Preliminaries - Wrapper Induction – Instance Based Wrapper Learning - Automatic Wrapper Generation: Problems - String Matching and Tree Matching - Multiple Alignment - Building DOM Trees - Extraction Based on a Single List Page and Multiple Pages.

UNIT V : Web Usage Mining

Web Usage Mining: Data Collection and Pre-Processing - Data Modeling for Web Usage Mining - Discovery and Analysis of Web Usage Patterns – Applications - Collaborative Filtering - Recommender Systems – Web Recommender systems based on User and Item

Course Outcome:

- Build a sample search engine using available open source tools.
- Identify the different components of a web page that can be used for mining.
- Apply machine learning concepts to web content mining.
- Implement Page Ranking algorithm and modify the algorithm for mining information.
- Process data using the Map Reduce paradigm.
- Design a system to harvest information available on the web to build recommender systems.
- Analyze social media data using appropriate data/web mining techniques.
- Modify an existing search engine to make it personalized.

TEXT BOOK:

1. 1.Bing Liu, " Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data (Data - Centric Systems and Applications)", Springer; 2nd Edition 2009

REFERENCE BOOKS:

1. Guandong Xu, Yanchun Zhang, Lin Li , "Web Mining and Social Networking: Techniques and Applications", Springer; 1st Edition .2010

2. Zdravko Markov, Daniel T. Larose, "Data Mining the Web: Uncovering patterns in Web Content, Structure, and Usage",John Wiley & Sons, Inc., 2007

3. Soumen Chakrabarti , "Mining the Web: Discovering Knowledge from Hypertext Data", Morgan Kaufmann; edition 2002