

Year: B. Tech III (Semester VI)

Subject Name: Web Intelligence and Mining
Type of course: Professional Elective Course

Subject Code: BTAI14604

Rationale: The world wide web is the largest information source that is easily accessible and searchable. The purpose of this subject is to provide an in-depth understanding of various aspects of mining the web, including indexing of documents on the web, algorithms to retrieve relevant information from the web and challenges to mining the web. It also introduces the student to useful applications of web mining which have impacted all aspects of our lives today.

Teaching and Examination Scheme:

Teaching Scheme				Theory Marks			Practical Marks		Total
L	T	P	C	TEE	CA1	CA2	TEP	CA3	
3	0	2	4	60	25	15	30	20	150

CA1: Continuous Assessment (assignments / projects / open book tests / closed book tests) CA2: Sincerity in attending classes / class tests / timely submissions of assignments / self-learning attitude / solving advanced problems TEE: Term End Examination TEP: Term End Practical Exam (Performance and viva on practical skills learned in course) CA3: Regular submission of Lab work / Quality of work submitted / Active participation in lab sessions / viva on practical skills learned in course.

Contents:

Sr. No.	Contents	Total Hrs
1.	Introduction to Web Mining: Motivation for Web Mining, Basic concepts of Information Retrieval, Information Retrieval Models-Boolean Model, Vector-Space Model, Statistical Language Model, Intelligent methods for Relevance Feedback, Issues in Web Mining.	10
2.	Web Page Pre-Processing, Indexing: Stopword Removal, Stemming, Web-page pre-processing, Detection of duplication; Inverted Index and its Compression, Latent Semantic Indexing;	06
3.	Web Page Ranking and Search: Algorithms for ranking- PageRank Algorithm, HITS; Web Search- Using similarity scores, Using Rank positions; combining multiple rankings; Web Search Engines.	06
4.	Web Crawling: Types of crawlers- Breadth first crawlers, Preferential crawlers, Universal crawlers, focused crawlers, topical crawlers; Implementation issues in crawling, Crawler ethics and conflicts	06

5.	Web Usage Mining- Sources and types of data, Data modelling for web usage mining, discovery and analysis of web usage patterns, Discovery and Analysis of Web Usage Patterns- session and visitor analysis, Cluster analysis and Visitor Segmentation, Analysis of Sequential and Navigation patterns, Classification and prediction based on web user transactions, Recommender Systems, Collaborative filtering- Filtering using K-NN, Filtering using Association Rules, Query Log Mining	10
6.	Applications of Web Mining: Web Spamming- Content Spamming, Link Spamming, Hiding techniques, Combating Spam; Social Network Analysis and Community Discovery; Mining Opinion from the Web; Business Applications of Web Mining	07

Suggested Specification table with Marks (Theory): (For B. Tech only)

Distribution of Theory Marks					
R Level	U Level	A Level	N Level	E Level	C Level
20	20	10	5	5	--

Legends: R: Remembrance; U: Understanding; A: Application, N: Analyze and E: Evaluate C: Create (Bloom’s Taxonomy)

Reference Books:

Sr. No	Title of book /article	Author(s)	Publisher and details like ISBN
1.	Data Mining: Concepts and Techniques	J. Han, M. Kamber	Morgan Kaufmann
2.	Web Data Mining: Exploring Hyperlinks, Contents and Usage Data	Bing Liu	Springer
3.	Mining the Web: Discovering Knowledge from Hypertext Data	Soumen Chakrabarti	Morgan Kaufmann Publishers

Note: Students should refer to the latest editions of books

Course Outcomes (CO):

Sr. No.	CO statements	Marks % weightage
CO-1	Explain the basic concepts of web content, structure and usage mining.	30%
CO-2	Analyze and discuss methods for web-page ranking and search	25%
CO-3	Select and apply suitable algorithms to mine patterns from web data	25%
CO-4	Implement various applications of intelligent web mining	20%

SARVAJANIK UNIVERSITY
Sarvajani College of Engineering and
Technology
Bachelor of Technology



List of Experiments:

1. Write a program to remove stopwords from an input text document. Note the compression achieved
2. Implement a program to classify text
3. Write a program to simulate a web crawler
4. Write a program to form clusters of consumers based on their buying patterns
5. Write a program to simulate a recommender system for an online shopping site