



# GUJARAT TECHNOLOGICAL UNIVERSITY

## BACHELOR OF ENGINEERING SYLLABUS

Minor Degree : DATA SCIENCE

Subject Code : 116AT02

Subject Name : Web Data Mining

WEF Academic Year :	2022 - 23
Semester :	6
Category of the Course :	Compulsory

### Course Objective :

- To learn how to extract data from the Web.
- To understand how to analyze collected data to derive the most information.

### Course Scheme :

Teaching Scheme			Total Credits	Assessment Pattern and Marks				Total Marks
L	T	PR	C	Theory		Practical		
				ESE (E)	PA(M)	ESE (V)	PA (I)	
3	0	0	3	70	0	0	0	70

### Course Content :

Sr. No.	Course Content	No. of Hours
1	<b>Module 1 :</b> Introduction to internet and WWW, Data Mining Foundations, Association Rules and Sequential Patterns, Basic Concepts of Association Rules, Apriori Algorithm, Frequent Itemset Generation, Association Rule Generation, Data Formats for Association Rule Mining, Mining with multiple minimum supports, Extended Model, Mining Algorithm, Rule Generation.	6
2	<b>Module 2 :</b> Mining Class Association Rules, Basic Concepts of Sequential Patterns, Mining Sequential Patterns on GSP, Mining Sequential Patterns on Prefix Span, Generating Rules from Sequential Patterns.	8
3	<b>Module 3 :</b> Concepts of Information Retrieval, IR Methods, Boolean Model, Vector Space Model and Statistical Language Model, Relevance Feedback, Evaluation Measures, Text and Web Page Pre-processing, Stopword Removal, Stemming, Web Page Preprocessing, Duplicate Detection, Inverted Index and Its Compression, Inverted Index, Search using Inverted Index, Index Construction, Index Compression, Latent Semantic Indexing, Singular Value Decomposition, Query and Retrieval, Web Search, Meta Search, Web Spamming.	10



# GUJARAT TECHNOLOGICAL UNIVERSITY

## BACHELOR OF ENGINEERING SYLLABUS

Minor Degree : DATA SCIENCE

Subject Code : 116AT02

Subject Name : Web Data Mining

4	<b>Module 4 :</b> Link Analysis, Social Network Analysis, Co-Citation and Bibliographic Coupling, Page Rank Algorithm, HITS Algorithm, CommModuley Discovery, Problem Definition, Bipartite Core CommModuleies, Maximum Flow CommModuleies, Email CommModuleies, Web Crawling, A Basic Crawler Algorithm – Breadth First Crawlers, Preferential Crawlers, Implementation Issues – Fetching, Parsing, Stopword Removal, Link Extraction, Spider Traps, Page Repository, Universal Crawlers, Focused Crawlers, Topical Crawlers, Crawler Ethics and Conflicts.	10
5	<b>Module 5 :</b> Opinion Mining, Sentiment Classification, Classification based on Sentiment Phrases, Classification Using Text Classification Methods, Feature based Opinion Mining and Summarization, Problem Definition, Object feature extraction, Comparative Sentence and 44 Relation Mining, Opinion Search and Opinion Spam. Web Usage Mining, Data Collection and Preprocessing, Sources and Types of Data, Key Elements of Web Usage Data Preprocessing, Data Modeling for Web Usage Mining, Discovery and Analysis of Web Usage Patterns, Session and Visitor Analysis, Cluster Analysis and Visitor Segmentation, Association and Correlation Analysis, Analysis of Sequential and Navigation Patterns.	8
<b>Total</b>		<b>42</b>

### References / Text Book :

1. Mining the Web: Discovering Knowledge from Hypertext Data, Soumen Chakrabarti, Morgan Kaufmann Publishers.
2. Bing Liu, Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data, Springer Publications, 2011.
3. Jiawei Han, Micheline Kamber, Data Mining: Concepts and Techniques, Second Edition, Elsevier Publications 2010.
4. Anthony Scime, Web Mining: Applications and Techniques, 2005.
5. Kowalski, Gerald, Mark T Maybury: Information Retrieval Systems: Theory and Implementation, Kluwer Academic Press, 1997.
6. Mathew Russell, Mining the Social Web 2nd Edition, Shroff Publisher/O'Reilly Publisher Publication.
7. Data Mining and Data Warehousing Principles and Practical Techniques, Parteek Bhatia, Cambridge University Press.



# GUJARAT TECHNOLOGICAL UNIVERSITY

## BACHELOR OF ENGINEERING SYLLABUS

Minor Degree : DATA SCIENCE

Subject Code : 116AT02

Subject Name : Web Data Mining

---

### Course Outcome :

After Completion of the Course, Student will able to :

No.	Course Outcomes
01	To explain how data is can be collected from the Web.
02	To extract data and information from the webpages.
03	To make decision based on the data collected.

\* \* \* \* \*