

# GUJARAT TECHNOLOGICAL UNIVERSITY

## COMPUTER ENGINEERING WEB DATA MANAGEMENT SUBJECT CODE:2180713 B.E. 8<sup>th</sup> SEMESTER

**Type of course:** Bachelor of Engineering

**Prerequisite:** Nil

**Rationale:** This course provides an in depth study of the area of web data management covering XML, XLink, and XPointer. The course primarily covers the state of the art in designing and building web applications and services, primarily focusing on issues and challenges that revolve around the management and processing of data. It also highlights the concepts of Ontology, RDF and OWL. It covers Building Web scale applications using web search, web crawlers, Web Graph mining, Map Reduce etc.

### Teaching and Examination Scheme:

Teaching Scheme			Credits	Examination Marks						Total Marks
L	T	P		Theory Marks			Practical Marks			
			ESE (E)	PA (M)		ESE (V)		PA (I)		
				PA	ALA	ESE	OEP			
3	0	2	5	70	20	10	20	10	20	150

### Content:

Sr. No.	Content	Total Hrs	% Weightage
1	<b>Data Model</b> Introduction to Modeling Web Data, Semistructured data, XML, Web Data Management with XML, XML Standards, XML and syntax, XML Data Model, XLink, and XPointer.	05	10
2	<b>XPath and XQuery</b> Introduction, Basics of XPath and XQuery, XPath: Steps and path expressions, Evaluation of path expressions, Generalities on axes and node tests, Axes, Node tests and abbreviations, Predicates, XPath 2.0; FLWOR expressions in XQuery: Defining variables - the for and let clauses, Filtering - the where clause, The return clause, Advanced features of XQuery; XPath foundations.	06	15
3	<b>Typing</b> Motivating Typing, Automata, Schema Languages for XML, Typing Graph Data: Graph Semistructured Data, Graph Bisimulation, Data guides.	04	10
4	<b>XML Query Evaluation</b> XML fragmentation, XML identifiers: Region-based identifiers, Dewey-based identifiers, Structural identifiers and updates; XML evaluation techniques: Structural join, Optimizing structural join queries, Holistic twig joins.	04	10
5	<b>Ontologies, RDF, and OWL</b>	05	15

	Introduction, Ontologies by example, Web resources, URI, namespaces, RDF, RDFS: RDF Schema, OWL, Ontologies and (Description) Logics.		
6	<b>Querying Data through Ontologies</b> Introduction, Querying RDF data: notation and semantics, Querying through RDFS ontologies, Answering queries through DL-LITE ontologies.	<b>04</b>	10
7	<b>Data Integration</b> Introduction, Containment of conjunctive queries, Global-as-view mediation, Local-as-view mediation, Ontology-based mediators, Peer-to-Peer Data Management Systems.	<b>05</b>	10
8	<b>Building Web scale applications</b> Web search, web crawlers, web information retrieval, Web graph mining and hot topics in web search, Distributed systems, failure management, Required properties of a distributed system, P2P networks, Hash-based structures, distributed indexing, Distributed computing with MapReduce.	<b>07</b>	20

**Suggested Specification table with Marks (Theory):**

Distribution of Theory Marks				
R Level	U Level	A Level	N Level	E Level
10	18	18	18	06

**Legends: R : Remembrance ; U = Understanding; A = Application; N = Analyze; E = Evaluation and above Levels (Revised Bloom's Taxonomy)**

**Reference Books:**

- Serge Abiteboul, Ioana Manolescu, Philippe Rigaux, Marie-Christine Rousset and Pierre Senellart, "Web Data Management", Cambridge University Press, 2011
- Bhavani Thuraisingham, "Web Data Management and Electronic Commerce", CRC Press, 2000
- Bhavani Thuraisingham, "XML Databases and the Semantic Web", CRC Press, 2002
- Athena Vakali and George Pallis, "Web Data Management Practices: Emerging Techniques and Technologies", IGI Publishing, 2007, ISBN-10: 1599042282; ISBN-13: 978-1599042282

**Course Outcome:**

After learning the course, the student will be able:

1. To understand the overall vision of the Semantic Web
2. To analyze the current technology stack (URIs, XML, RDF/S, OWL)
3. To understand how one could use these technologies for building something useful
4. To define and test an ontology
5. To define schema mappings
6. To install and use tools for semantic data management

**List of Experiments:**

1. Create an XML file defining an article in newspaper.
2. Create an XML file containing list of students. Also create stylesheet file to display list in an HTML format.
3. Create an XML file containing list of students. Using XPath display following information
  - Information of a student with ID No : 101
  - All the student in the sorted order according to their CGPA
4. Create an XForm to collect information from staff member regarding their publications. Details like Year of Publication, National/International, Title, Conference/Journal etc.
5. From the above gathered information, using XQuery find out the number of publication in a specific year.
6. Demonstrate the use of AJAX.
7. Study of XMLSPY tool.
8. Create an RSS for the events occurring in your institute.
9. Write a program to read the articles in RSS created in above practical.
10. Study of RDF (Resource Description Framework)

**Design based Problems (DP)/Open Ended Problem:**

1. Create a system which can use of Web search, web crawlers and web information retrieval.
2. Analyze and implement a system with Web graph mining.
3. Implement and Subscribe RSS News feeds to get latest news in India.

**Major Equipment:**

XMLSpy, RSS Feed, RSS Reader.

**List of Open Source Software/learning website:**

- <http://in.reuters.com/tools/rss>
- <http://www.altova.com/xmlspy.html>
- <https://www.w3.org/RDF/>

**ACTIVE LEARNING ASSIGNMENTS:** Preparation of power-point slides: which include videos, animations, pictures, graphics for better understanding theory and practical work. The faculty will allocate chapters/ parts of chapters to groups of students so that the entire syllabus can be covered. The power-point slides should be put up on the web-site of the College/ Institute, along with the names of the students of the group, the name of the faculty, Department and College on the first slide. The best three works should submit to GTU.