Processing XML with Java

CDT757

This course builds the skills needed by developers and others who wish to create Java applications that process XML. Participants will learn to design code and test Java application programs that can be used in XML data interchange applications and in specialized XML publishing applications. Topics covered include the XML-RPC and SOAP XML-based protocols as well as the SAX, DOM and JDOM processing models.

Audience
- XML developers and systems architects who need to know how develop Java applications that process XML.

Prerequisites
- CDT753, Introduction to XML
- Java programming experience

Course Length
- Three days

Teaching Methods
- Lecture
- Hands-on examples
- Hands-on supplemental exercises

Learning Objectives
- Understand the suitability of Java for processing XML
- Understand the use of the XML-RPC and SOAP protocols
- Write a Java program that writes XML directly
- Write a Java program the converts a flat file to an XML file
- Write XML-RPC clients using SAX and DOM
- Write SOAP clients using SAX and DOM
- Write an XML processing program using SAX
- Write a SAX filter program
- Write a program that creates an XML document using DOM
- Write a DOM-based SOAP servlet
- Write a program that creates an XML document using JDOM
- Write a JDOM-based program that checks XML documents for well-formedness
- Write a JDOM-based program that checks XML documents for validity
- Understand the applicability of SAX, DOM and JDOM processing models to various XML processing tasks

Course Outline

TG5

XML for Data
- Motivating XML
- XML Syntax
- Validity
- Stylesheets

XML Protocols: XML-RPC and SOAP
- XML as a Message Format
- HTTP as a Transport Protocol
- RSS
- Customizing the Request
- XML-RPC
- SOAP

Writing XML with Java
- Fibonacci Numbers
- Writing XML
- Output Streams, Writers and Encodings
- A Simple XML-RPC Client
- A Simple SOAP Client
- Servlets

Converting Flat Files to XML
- The Budget
- The Model
- Input
- Determining the Output Format
- Building Hierarchical Structures from Flat Files
- Alternatives to Java
- Relational Databases

Reading XML
- InputStreams and Readers
- XML Parsers
- SAX
- DOM
- JAXP
- JDOM
- dom4

SAX
- What is SAX?
- Parsing
- Callback Interfaces
- Receiving Documents
- Receiving Elements
- Handling Attributes
- Receiving Characters
- Receiving Processing Instructions
- Receiving Namespace Mappings
- "Ignorable White Space"
- Receiving Skipped Entities
- Receiving Locators
- What the ContentHandler Doesn't Tell You

The XMLReader Interface
- Building Parser Objects
- Input
- Exceptions and Errors
- Features and Properties
- DTDHandler

Sax Filters
- The Filter Architecture
- The XMLFilter Interface
- Content Filters
- The XMLFilterImpl Class
- Parsing Non-XML documents
- Multihandler Adapters

The Document Object Model (DOM)
- The Evolution of DOM
- DOM Modules
- Application-Specific DOMs
- Trees
- DOM Parsers for Java
- Parsing Documents with a DOM Parser
- The Node Interface
- The NodeList Interface
- JAXP Serialization
- DOMException
- Choosing Between SAX and DOM

Creating XML Documents with DOM
- DOMImplementation
- Locating a DOMImplementation
- The Document Interface as an Abstract Factory
- The Document Interface as a Node Type
- Normalization

The DOM Core
- The Element Interface
- The NamedNodeMap Interface
- The CharacterData Interface
- The Text Interface
- The CDATASection Interface
- The EntityReference Interface
- The Attr Interface
- The ProcessingInstruction Interface
- The Comment Interface
- The DocumentType Interface
- The Entity Interface
- The Notation Interface

The DOM Traversal Module
- NodeIterator
- NodeFilter
- TreeWalker

Output from DOM
- Xerces Serialization
- OutputFormat
- DOM Level 3

JDOM
- What is JDOM?
- Creating XML Elements with JDOM
- Creating XML Documents with JDOM
- Writing XML Documents with JDOM
- Document Type Declarations
- Namespaces
- Reading XML Documents with JDOM
- Navigating JDOM Trees
- Talking to DOM Programs
- Talking to SAX Programs
- Java Integration
- What JDOM Doesn't Do

The JDOM Model
- The Document Class
- The Element Class
- The Attribute Class
- The Text Class
- The CDATA Class
- The ProcessingInstruction Class
- The Comment Class
- Namespaces
- The DocType Class
- The EntityRef Class