**XML** (Extensible Markup Language)

* a text-based markup language
* standard for data interchange on the web.
* to store all types of data in a structured form

Collectively, the tags are known as markup.

Purpose of markup languages

* provide a cross-platform tool for safe transmission of information in the form of texts, pictures etc. through internet.

Hyper-text

* A document that contains links to other documents (and text, sound, images...)

Markup language

* A notation for writing text with markup tags
* The tags indicate the structure of the text
* Tags have names and attributes
* Tags may enclose a part of the text

Unfortunately XML alone has no powers to display its contents just like HTML. So some supporting languages like DTD (Document Type Definition) for validation.

XSL (eXtensible Style Language) and CSS (Cascade Style Sheet) for purposes of processing and display were also developed along with XML’s.

**XPath** is used to navigate through elements and attributes in an XML document.

**XML Parsing** - reading the XML file/string and getting its content according to the structure, usually to use them in a program. (Here in my case it’s java program)

**JAXP** (Java API for XML Processing) - the Java XML Application programming interfaces (API).

Validating and parsing XML documents. The 3 basic parsing interfaces are:

* Document Object Model (**DOM**) parsing interface
* Simple API for XML parsing interface or **SAX** interface
* Streaming API for XML or **StAX** interface (part of JDK 6; separate jar available for JDK 5)

**JAXB** (Java API for XML Binding)

1. Writing out Java objects as XML (**marshalling**).
2. Creating Java objects from XML structures (**unmarshalling**).

xjc can be used to convert XML Schema to class.

Classes are marked up using annotations from javax.xml.bind.annotation.\* namespace, for example, @XmlRootElement and @XmlElement.

Marshallers and Unmarshallers are created through an instance of JAXBContext.

**JAX-WS** (Java API for XML Web Services) is a Java programming language API for creating web services.

JAX-WS is one of the Java XML programming APIs.

The JAX-WS 2.2 specification JSR 224 defines a **standard Java-to-WSDL mapping** which determines how WSDL operations are bound to Java methods when a SOAP message invokes a WSDL operation.

This Java-to-WSDL mapping determines which Java method gets invoked and how that SOAP message is mapped to the method’s parameters.

JAX-WS uses annotations, introduced in Java SE 5, to simplify the development and deployment of web service clients and endpoints.

JAX-WS 2.0 replaced the JAX-RPC API in Java Platform, Enterprise Edition 5 which leans more towards document style Web Services.

**JAX-RS** (Java API for RESTful Services)

Portable APIs for developing, exposing and accessing Web applications designed and implemented in compliance with principles of REST architectural style.

**Implementations of JAX-RS**:

**Apache CXF**, an open source Web service framework

Jersey, the reference implementation from Sun (now Oracle)

**Jersey** - Jersey RESTful Web Services framework is open source framework for developing RESTful Web Services in Java that provides support for JAX-RS APIs and serves as a JAX-RS (JSR 311 & JSR 339) Reference Implementation.

JAX-RS supports the creation of XML and JSON via the Java Architecture for XML Binding (JAXB).

**DTD (Document Type Definition)**

The purpose of a DTD is to define the legal building blocks of an XML document.

A DTD defines the document structure with a list of legal elements and attributes.

A DTD can be declared inline inside an XML document, or as an external reference.

Why Use DTD?

Use a standard DTD for interchanging data.

Use a standard DTD to verify the data you receive from the outside world is valid.

Use a DTD to verify your own data.

Legal building blocks

* Elements
* Attributes
* Entities
* PCDATA - parsed character data (PCDATA is text that WILL be parsed by a parser)
* CDATA -character data.(CDATA is text that will NOT be parsed by a parser)

DTD Validation

You can validate your XML against a DTD

The load( ) method is used for files, while the loadXML( ) method is used for strings.

Validation can be turned off by setting the XML parser's validateOnParse="false".

**XSD**

An XML Schema describes the structure of an XML document.

XML Schema is an XML-based alternative to DTD.

**DOM parser and SAX Parser in Java** (High level differences)

|  |  |
| --- | --- |
| DOM | SAX |
| Suitable for **small xml file** | Suitable for **large xml file** |
| **Loads the whole xml doc** in memory | SAX **only loads small part of XML** file in memory |
| Creates a **tree structure** in memory | Is **event based**, takes components in xml as events, does not create any structure in memory |

**DOM** presents an XML document as a tree-structure.

The DOM is a W3C (World Wide Web Consortium) standard.

The DOM is separated into 3 different parts / levels:

Core DOM - standard model for any structured document

XML DOM - standard model for XML documents

HTML DOM - standard model for HTML documents

The XML DOM defines a standard way for accessing and manipulating XML documents.

The XML DOM defines the objects and properties of all XML elements, and the methods (interface) to access them.

The XML DOM is a standard for how to get, change, add, or delete XML elements.

|  |  |
| --- | --- |
| HTTP | HTTPS |
| URL begins with “http://” | URL begins with “https://” |
| Port 80 for communication | Port 443 for communication |
| Unsecured | Secured |
| Operates at Application Layer | Operates at Transport Layer |
| No encryption | Encryption is present |
| No certificates required | Certificates required |

A communications protocol is a system of digital rules for data exchange within or between computers.

When data is exchanged through a computer network, the rules system is called a network protocol.

**HTTP (Hypertext Transfer Protocol)**

An application protocol

Is the foundation of data communication for the World Wide Web.

HTTP is the protocol to exchange or transfer hypertext.

Hypertext is structured text that uses logical links (hyperlinks) between nodes containing text.

HTTP functions as a request-response protocol in the client-server computing model.

An HTTP session is a sequence of network request-response transactions.

HTTP is a stateless protocol. A stateless protocol does not require the HTTP server to retain information or status about each user for the duration of multiple requests.

TCP/IP software is organized in four layers.

* Application layer
* Transport layer
* Internet layer
* Network interface layer.

High level interaction protocols are:

* IIOP
* RMI
* SOAP

Java Community Process (JCP) – Community development for Java technology specification.

Open source refers to software, and open standards refer to documents (that may then be implemented by software).