

DOM-Based XML Processing

- **DOM (Document Object Model):** A way of representing and interacting with XML as a tree structure in memory, where each element, attribute, and text is a node.
- **DOM Processing:** Allows manipulation of XML data with methods like `getElementById`, `getElementsByTagName`, etc.

XML DOM

What is XML DOM

DOM is an acronym stands for Document Object Model. It defines a standard way to access and manipulate documents. The Document Object Model (DOM) is a programming API for HTML and XML documents. It defines the logical structure of documents and the way a document is accessed and manipulated.

one important objective for the Document Object Model is to provide a standard programming interface that can be used in a wide variety of environments and applications. The Document Object Model can be used with any programming language.

XML DOM defines a standard way to access and manipulate XML documents.

What does XML DOM

The XML DOM makes a tree-structure view for an XML document.

We can access all elements through the DOM tree.

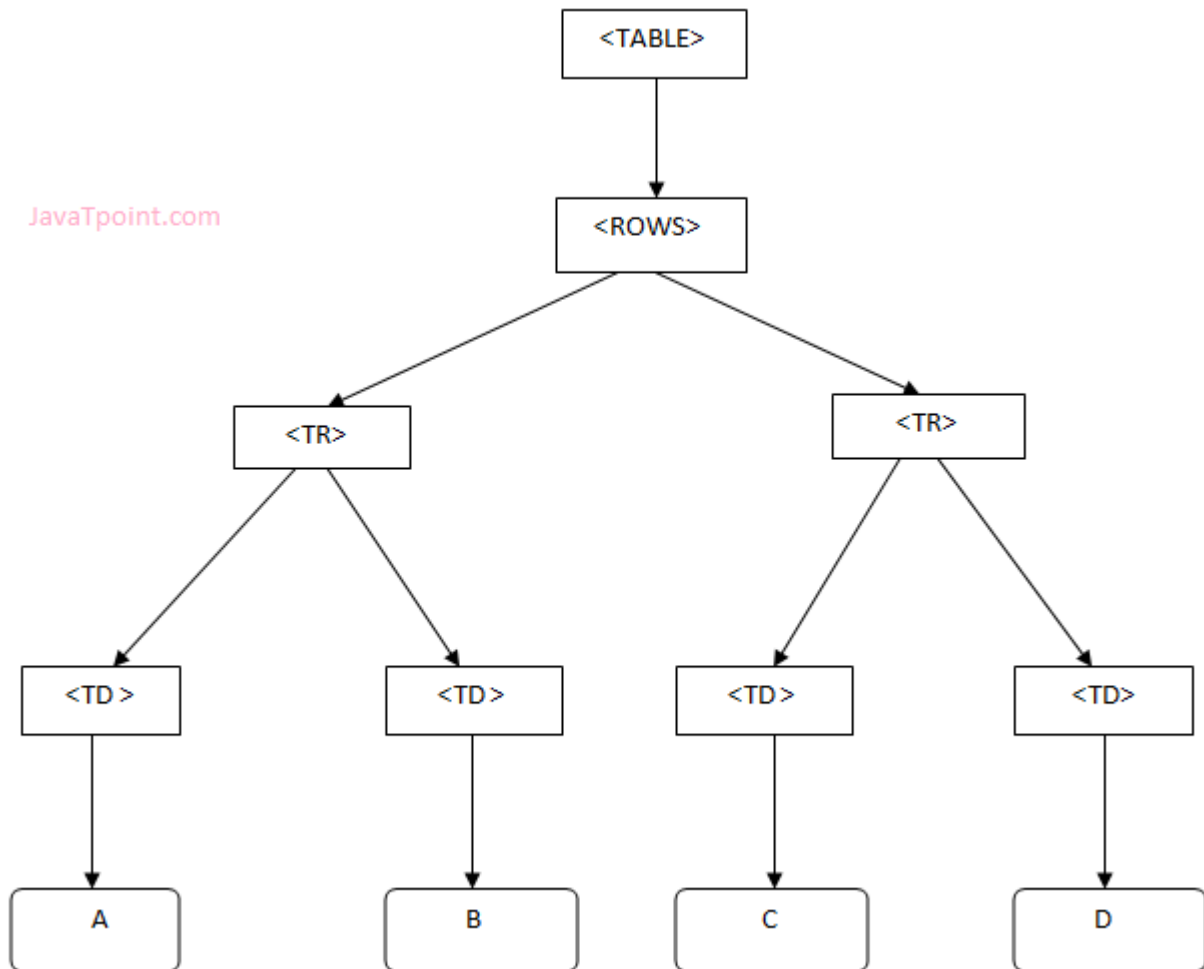
We can modify or delete their content and also create new elements. The elements, their content (text and attributes) are all known as nodes.

For example, consider this table, taken from an HTML document:

1. `<TABLE>`
2. `<ROWS>`
3. `<TR>`
4. `<TD>A</TD>`
5. `<TD>B</TD>`
6. `</TR>`
7. `<TR>`
8. `<TD>C</TD>`
9. `<TD>D</TD>`
10. `</TR>`
11. `</ROWS>`

12. </TABLE>

The Document Object Model represents this table like this:



Event-Oriented Parsing: SAX

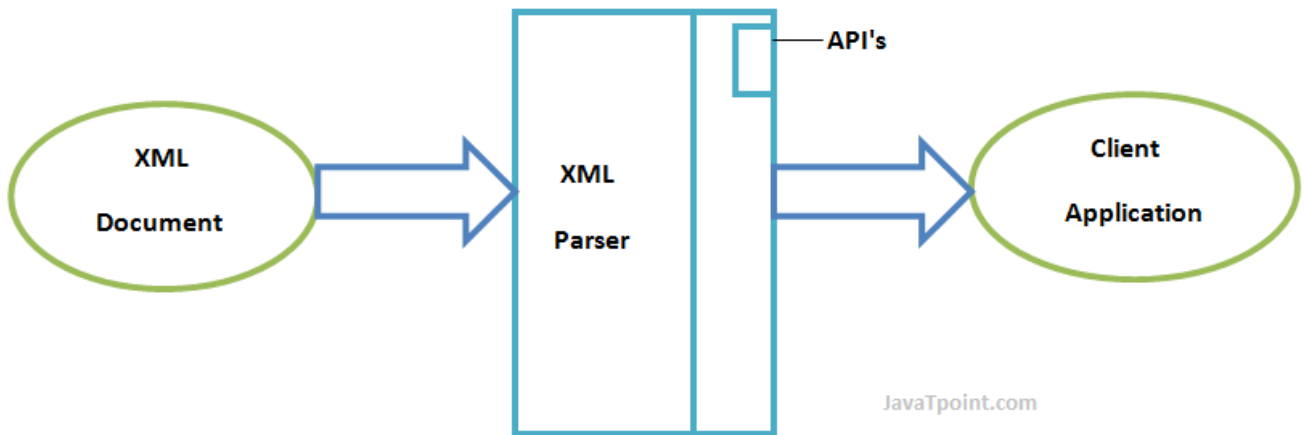
- **SAX (Simple API for XML):** An event-driven, serial-access mechanism for parsing XML. Unlike DOM, SAX does not load the entire XML document into memory, making it suitable for large documents.
- **SAX Parsing:** SAX parser reads the XML line by line, triggering events (start and end of elements, character data) which can be processed with callbacks.

XML Parsers

An XML parser is a software library or package that provides interfaces for client applications to work with an XML document. The XML Parser is designed to read the XML and create a way for programs to use XML.

XML parser validates the document and check that the document is well formatted.

Let's understand the working of XML parser by the figure given below:



Types of XML Parsers

These are the two main types of XML Parsers:

❑ DOM

❑ SAX

DOM (Document Object Model)

A DOM document is an object which contains all the information of an XML document. It is composed like a tree structure. The DOM Parser implements a DOM API. This API is very simple to use.

Features of DOM Parser

A DOM Parser creates an internal structure in memory which is a DOM document object and the client applications get information of the original XML document by invoking methods on this document object.

DOM Parser has a tree based structure.

Advantages

- 1) It supports both read and write operations and the API is very simple to use.
- 2) It is preferred when random access to widely separated parts of a document is required.

Disadvantages

- 1) It is memory inefficient. (consumes more memory because the whole XML document needs to be loaded into memory).
- 2) It is comparatively slower than other parsers.

SAX (Simple API for XML)

A SAX Parser implements SAX API. This API is an event based API and less intuitive.

Features of SAX Parser

It does not create any internal structure.

Clients does not know what methods to call, they just overrides the methods of the API and place his own code inside method.

It is an event based parser, it works like an event handler in Java.

Advantages

- 1) It is simple and memory efficient.
- 2) It is very fast and works for huge documents.

Advertisement

Disadvantages

- 1) It is event-based so its API is less intuitive.
 - 2) Clients never know the full information because the data is broken into pieces.
-
-

Transforming XML Documents

- **XSLT (Extensible Stylesheet Language Transformations):** A language for transforming XML documents into other formats (e.g., HTML, plain text).
- **Transformation Process:** XSLT uses templates that match specific XML elements and defines how each element should be transformed.

XSL stands for EXtensible Stylesheet Language. XSLT is for Transformation of XML document to other formats.

What is XSLT

Before XSLT, first we should learn about XSL. XSL stands for EXtensible Stylesheet Language. It is a styling language for XML just like CSS is a styling language for HTML.

XSLT stands for XSL Transformation. It is used to transform XML documents into other formats (like transforming XML into HTML).

What is XSL

In HTML documents, tags are predefined but in XML documents, tags are not predefined. World Wide Web Consortium (W3C) developed XSL to understand and style an XML document, which can act as XML based Stylesheet Language.

An XSL document specifies how a browser should render an XML document.

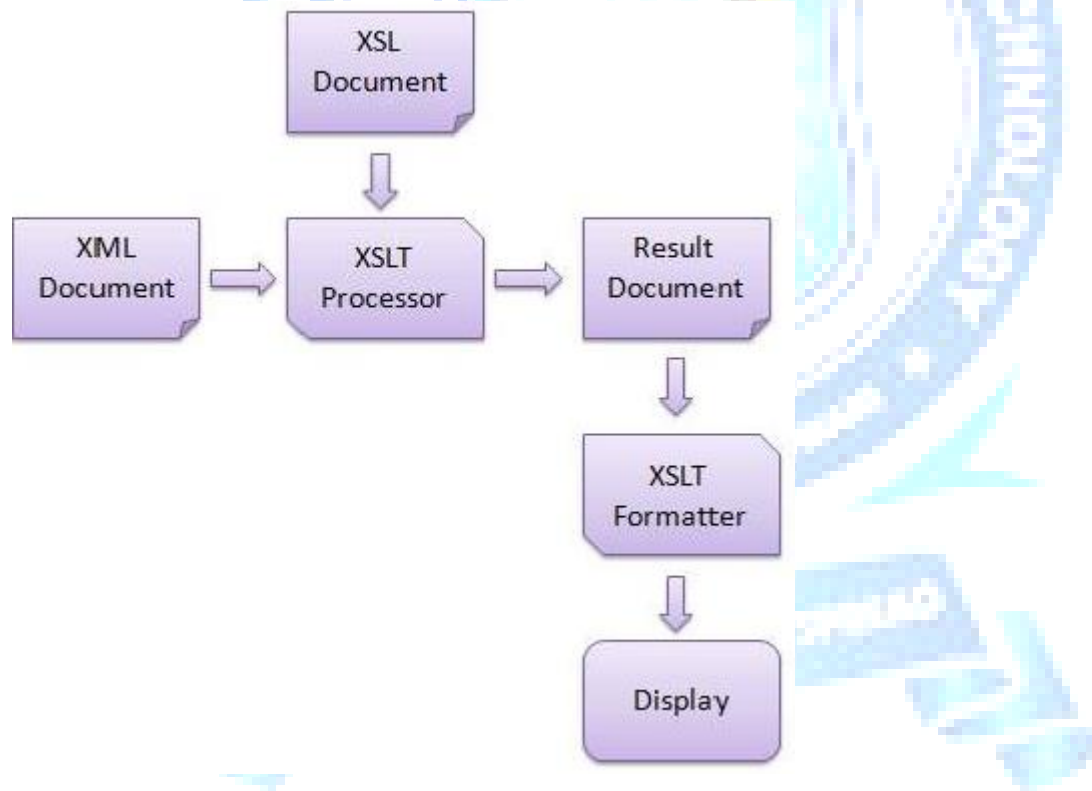
Main parts of XSL Document

- **XSLT:** It is a language for transforming XML documents into various other types of documents.
- **XPath:** It is a language for navigating in XML documents.
- **XQuery:** It is a language for querying XML documents.
- **XSL-FO:** It is a language for formatting XML documents.

How XSLT Works

The XSLT stylesheet is written in XML format. It is used to define the transformation rules to be applied on the target XML document. The XSLT processor takes the XSLT stylesheet and applies the transformation rules on the target XML document and then it generates a formatted document in the form of XML, HTML, or text format. At the end it is used by XSLT formatter to generate the actual output and displayed on the end-user.

Image representation:



Advantage of XSLT

A list of advantages of using XSLT:

- XSLT provides an easy way to merge XML data into presentation because it applies user defined transformations to an XML document and the output can be HTML, XML, or any other structured document.

- XSLT provides XPath to locate elements/attribute within an XML document. So it is more convenient way to traverse an XML document rather than a traditional way, by using scripting language.
- XSLT is template based. So it is more resilient to changes in documents than low level DOM and SAX.
- By using XML and XSLT, the application UI script will look clean and will be easier to maintain.
- XSLT templates are based on XPath pattern which is very powerful in terms of performance to process the XML document.
- XSLT can be used as a validation language as it uses tree-pattern-matching approach.
- You can change the output simply modifying the transformations in XSL files.

Selecting XML Data: XPath

- **XPath (XML Path Language):** A query language used to select nodes or data from an XML document.
- **Usage of XPath:** Allows accessing specific elements, attributes, and portions of the document using path expressions. Example:

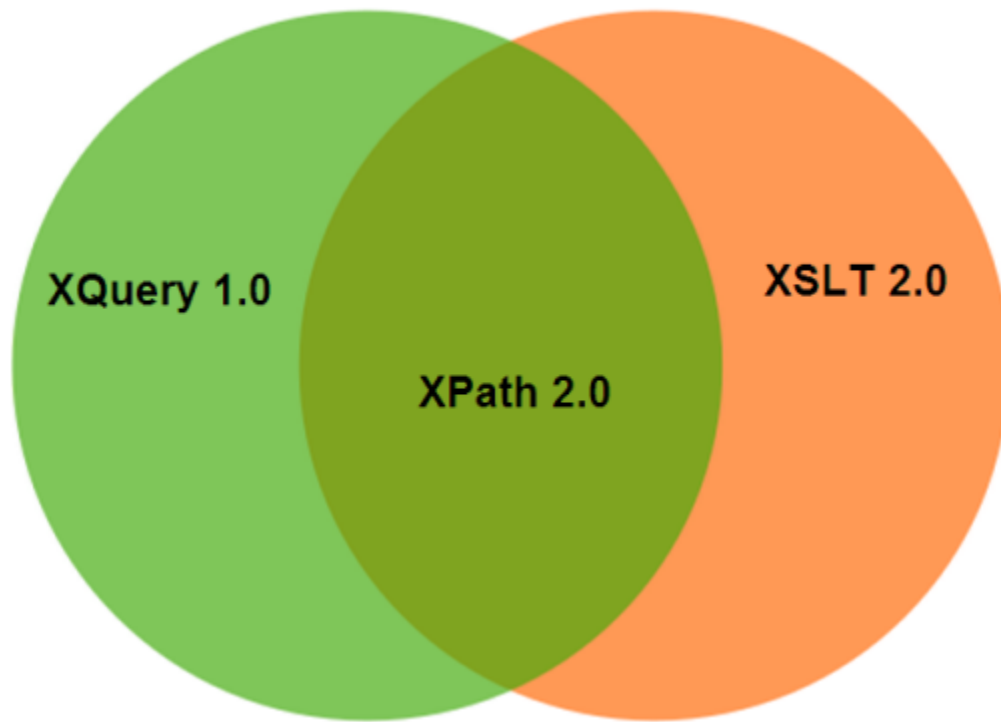
xpath

/bookstore/book[price>30]

What is XPath

XPath is an important and core component of XSLT standard. It is used to traverse the elements and attributes in an XML document.

XPath is a W3C recommendation. XPath provides different types of expressions to retrieve relevant information from the XML document. It is syntax for defining parts of an XML document.



Important Features of XPath

- **XPath defines structure:** XPath is used to define the parts of an XML document i.e. element, attributes, text, namespace, processing-instruction, comment, and document nodes.
- **XPath provides path expression:** XPath provides powerful path expressions, select nodes, or list of nodes in XML documents.
- **XPath is a core component of XSLT:** XPath is a major element in XSLT standard and must be followed to work with XSLT documents.
- **XPath is a standard function:** XPath provides a rich library of standard functions to manipulate string values, numeric values, date and time comparison, node and QName manipulation, sequence manipulation, Boolean values etc.
- **Path is W3C recommendation.**