* XML is abbreviation for eXtensible Markup Language whereas HTML stands for Hypertext Markup Language.
* XML mainly focuses on transfer of data while HTML is focused on presentation of the data.
* XML is content driven
* **XML is Case sensitive** while HTML is Case not sensitive.
* XML provides namespaces support while HTML doesn't provide namespaces support.
* XML is strict for closing tag while HTML is not strict.
* XML tags are extensible whereas HTML has limited tags.
* XML tags are not predefined whereas HTML has predefined tags.

**As SQL is used to retrieve data from database, XPath is used to retrieve data from XML.**

**DOM**:

DOM is the acronym for Document Object Model. It’s a cross platform and language neutral standard that defines how to access and manipulate data in;

* HTML
* XHTML
* XML

DOM XML is used to access and manipulate XML documents. **It views the XML document as a tree-structure.**

An XML schema is used to define the structure of an XML document. It is like DTD but provides more control on XML structure.

**DTD (Document Type Definition) and XSD (XML Schema Definition) are used to define XML structure.**

**XML Parsers**

An XML parser is a program that translates the **XML document** into an **XML Document Object Model (DOM) Object.**

The XML DOM Object can then be manipulated using JavaScript, Python, and PHP etc.

The keyword CDATA which is the acronym for (Unparsed) Character Data is used to ignore special characters such as “<,>” when parsing an XML document.

**CDATA vs PCDATA**

CDATA: (Unparsed Character data): CDATA contains the text which is not parsed further in an XML document. Tags inside the CDATA text are not treated as markup and entities will not be expanded.

Let's take an example for CDATA:

**<?xml** version="1.0"**?>**

<!DOCTYPE employee SYSTEM "employee.dtd"**>**

**<employee>**

<![CDATA[

  <firstname>vimal</firstname>

  <lastname>jaiswal</lastname>

  <email>vimal@javatpoint.com</email>

]]>

**</employee>**

Output: <firstname>vimal</firstname><lastname>jaiswal</lastname><email>vimal@javatpoint.com</email>

PCDATA: (Parsed Character Data): XML parsers are used to parse all the text in an XML document. PCDATA stands for Parsed Character data.

**<?xml** version="1.0"**?>**

<!DOCTYPE employee SYSTEM "employee.dtd"**>**

**<employee>**

**<firstname>**vimal**</firstname>**

**<lastname>**jaiswal**</lastname>**

**<email>**vimal@javatpoint.com**</email>**

**</employee>**

Output: vimal jaiswal vimal@javatpoint.com

**Web** **services** such as SOAP and REST use XML format to exchange information.

## XML Validation

XML file can be validated by 2 ways:

1. against DTD
2. against XSD

DTD (Document Type Definition) and XSD (XML Schema Definition) are used to define XML structure.