

---

# Introduction to XML

Instructor: Aarthi Natarajan

April 26<sup>th</sup>, 2017

---

# Objectives

- XML stands for eXtensible Markup Language – a text based markup language
- XML is used for storing and transporting data through the use of XML tags
- XML does not replace HTML, but complements HTML
- XML is a public standard, developed by the W3C (World Wide Web Consortium)
- XML is just information wrapped in tags, XML itself does not do anything
- Someone must develop software to send, receive, store and display it

# Difference between XML and HTML

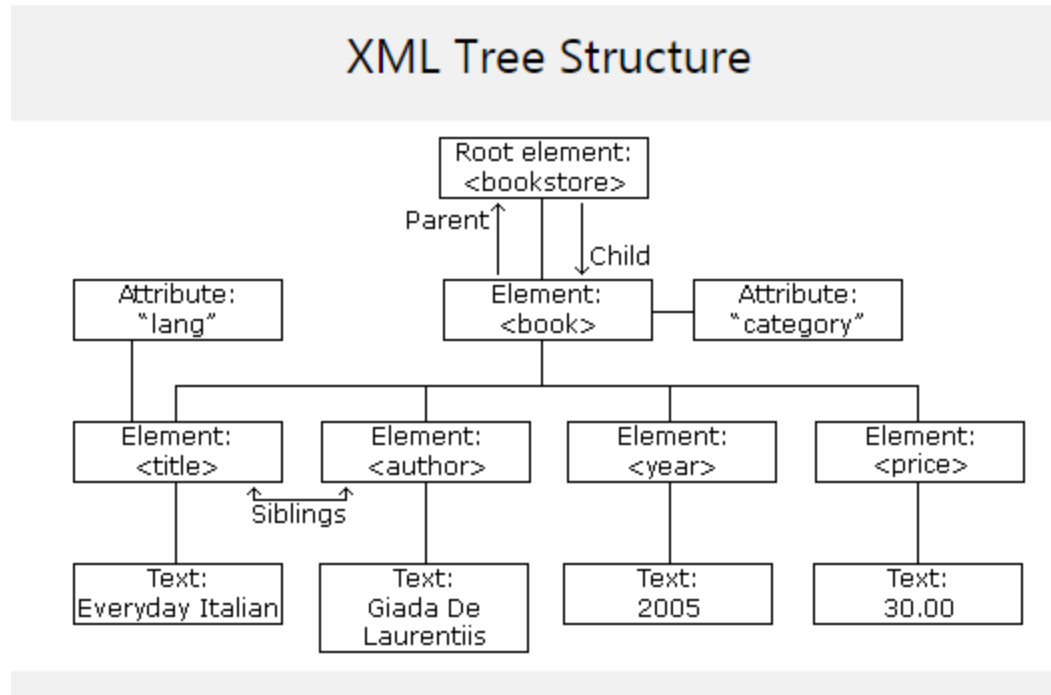
<p>XML is a markup language that was designed to carry data – with a focus on what the data is. XML is about describing information</p>	<p>HTML is a markup language that was designed to display data (web pages) in a browser – with a focus on what the data looks like. HTML is about displaying data</p>
<p>Dynamic</p>	<p>Static</p>
<p>XML does not use pre-defined tags. The author of the XML file “invents” the tags to be used. The author can specify the structure of the XML document or allowable set of tags through defining an XML schema</p>	<p>HTML makes use of predefined tags such as &lt;p&gt;, &lt;h1&gt;</p>
<p>XML stores data in plain-text, to provide a software and hardware independent way of storing, transporting and sharing data</p>	<p>HTML also stores data in plain-text</p>
<p>XML can be used to store data in files and databases. As XML separates data from presentation, the same XML file can be used to display in different presentation scenarios</p>	<p>HTML pages are purely used to display data.</p>

# Sample XML file: Example1

```
<note>  
  <to>Larry</to>  
  <from>Joe</from>  
  <heading>Reminder</heading>  
  <body>Do not forget the CSE school meeting</body>  
</note>
```

# XML Tree

XML documents form a tree structure that starts at the “root” and branches to the “leaves”.



# XML Example 2

The image in the previous slide is represented as an XML document as below:

```
<?xml version="1.0" encoding="UTF-8"?>
<bookstore>
  <book category="cooking">
    <title lang="en">Everyday Italian</title>
    <author>Giada De Laurentiis</author>
    <year>2005</year>
    <price>30.00</price>
  </book>
  <book category="children">
    <title lang="en">Harry Potter</title>
    <author>J K. Rowling</author>
    <year>2005</year>
    <price>29.99</price>
  </book>
  <book category="web">
    <title lang="en">Learning XML</title>
    <author>Erik T. Ray</author>
    <year>2003</year>
    <price>39.95</price>
  </book>
</bookstore>
```

# XML Element

- The building blocks of XML documents
- Syntax of an element is:

**<element-name attribute1 attribute2> ....content </element-name**

where

element-name is the name of the element and both the start and end tags of the element must be identical (in case as well)

attribute1, attribute2 are properties of the element and associates a property name to a value. The value must be in quotes. e.g:

**<book category="fiction"> .....**

- An element is a container, that can contain text or other elements as seen in the previous example.

# XML Syntax Rules

- Every XML document must have a root element, which is the parent of all other elements
- An XML document can have an optional prolog, which if present, must be the first line of the XML document

```
<?xml version="1.0" encoding="UTF-8"?>
```

- All XML elements must have a closing tag e.g:

```
<book> Some book </book>
```

- XML tags are case-sensitive, opening and closing-tags must be written in the same case and must match
- XML elements must be properly nested
- XML attribute values must be quoted
- An XML document that adheres to all the syntax rules is said to be a **“well-formed”** XML document
- An empty element has the following syntax - `<element_name attr1 attr2 ../>`



# XML Schema

- The structure of an XML document is described by an XML Schema

An example of an XML schema for the sample XML Example 1:

```
<xs:element name="note">

<xs:complexType>
  <xs:sequence>
    <xs:element name="to" type="xs:string"/>
    <xs:element name="from" type="xs:string"/>
    <xs:element name="heading" type="xs:string"/>
    <xs:element name="body" type="xs:string"/>
  </xs:sequence>
</xs:complexType>

</xs:element>
```

- An XML document validated against an XML Schema is both "**Well Formed**" and "**Valid**".