Introduction to XML, Part 1
Building a Website with HTML

Getting started:
1) Create a directory (folder) in a convenient place on your computer and call it XML-workshop.

2) Download the following files from the location on the board and save them to your XML-workshop folder.

pirates-instructions.pdf
template.html
style.css
info.txt
The CaseFiles directory and all of its contents
wellformed-sheet.pdf

3) Download and install the Atom text editor. Installation instructions can be found here: https://atom.io/docs/v0.194.0/getting-started-installing-atom

DLab April 19 & 26, 2016
Scott Paul McGinnis

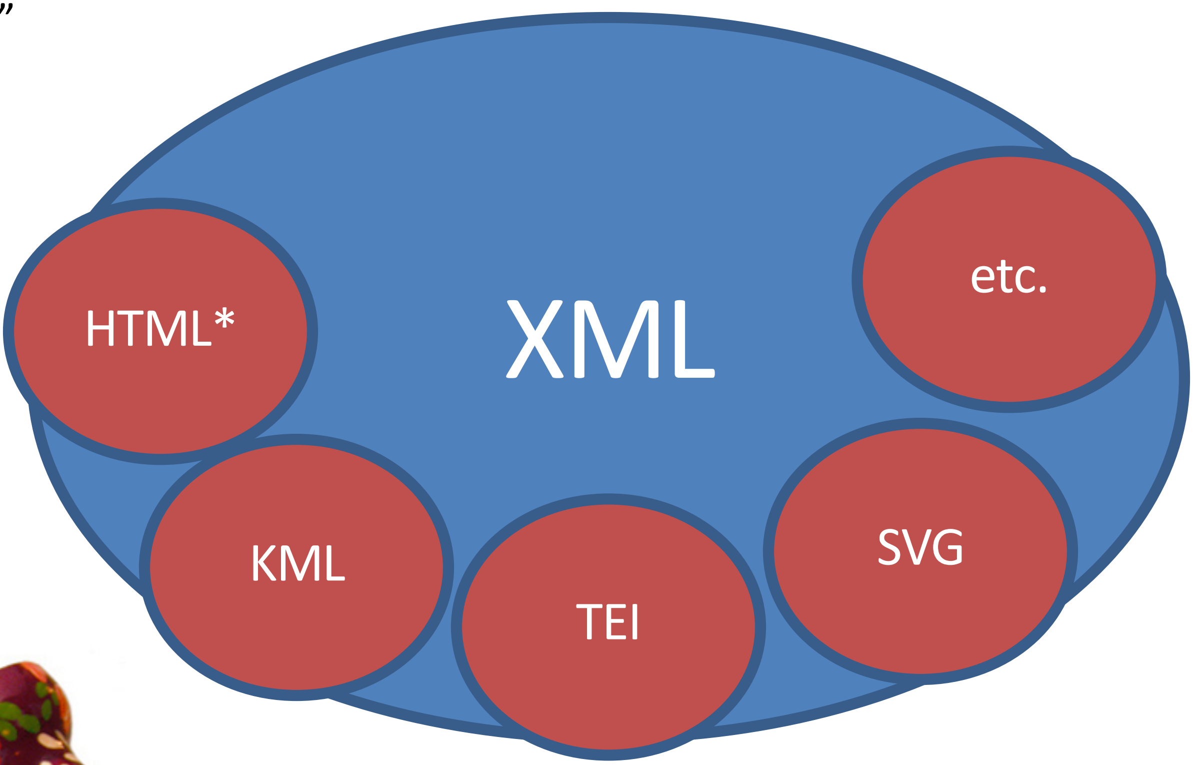
What is XML?

“eXtensible Markup Language”

- A markup language

- Like a language family (extensibility)

- Human readable and fairly easy to create



Some XML Applications

Should I learn XML?

Advantages
Ubiquity
Relative Ease
Made for Text
Helpful with Web-scraping / APIs
Well-maintained by a large and active community
Unicode Compliant

Helpful XML Principles
Extensible
Human Readable*
Separation of Content and Display

But... XML doesn't do very much by itself...
Websites: HTML + CSS + Javascript/JQuery
Wordpress: HTML + CSS + PHP
XRX databases: XML + Restful APIs + xQuery
Web-scraping: HTML + scripting (e.g. Python)
APIs: XML + URL queries + scripting

Examples in Projects and Research

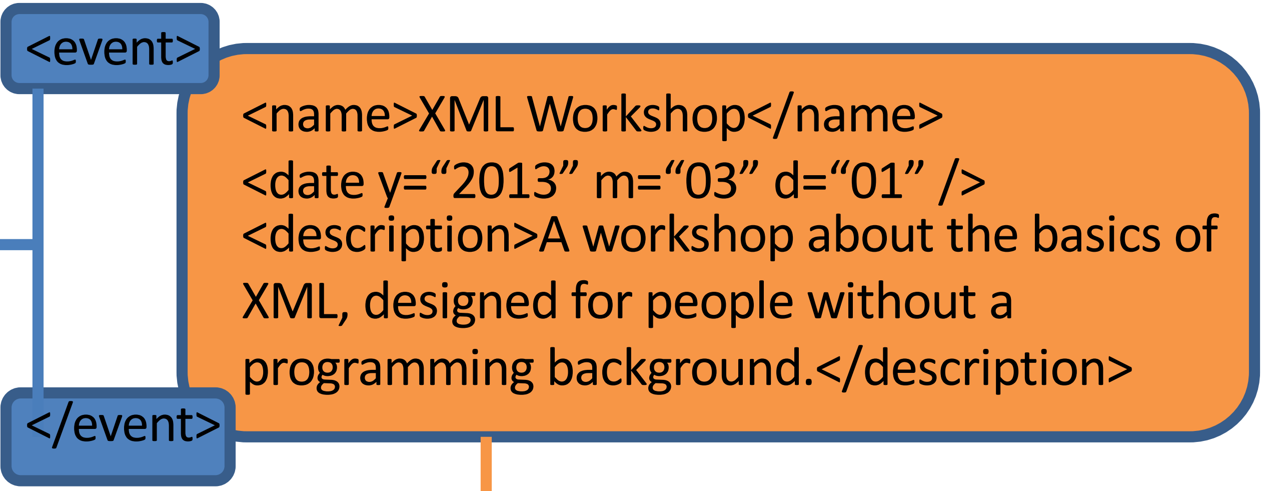
(switch to the other slides)

XML Basic Concepts: Elements and Content

The basic unit of XML is called an Element. Think of it like a container for your data or text that describes and categorizes it for the computer. When viewing a web page, for example, you will not see the elements. But you can access and manipulate the elements in all sorts of ways.

This is an Element. It has two components: the Start Tag and the End Tag.

Notice the use and order of the special characters:
< > /

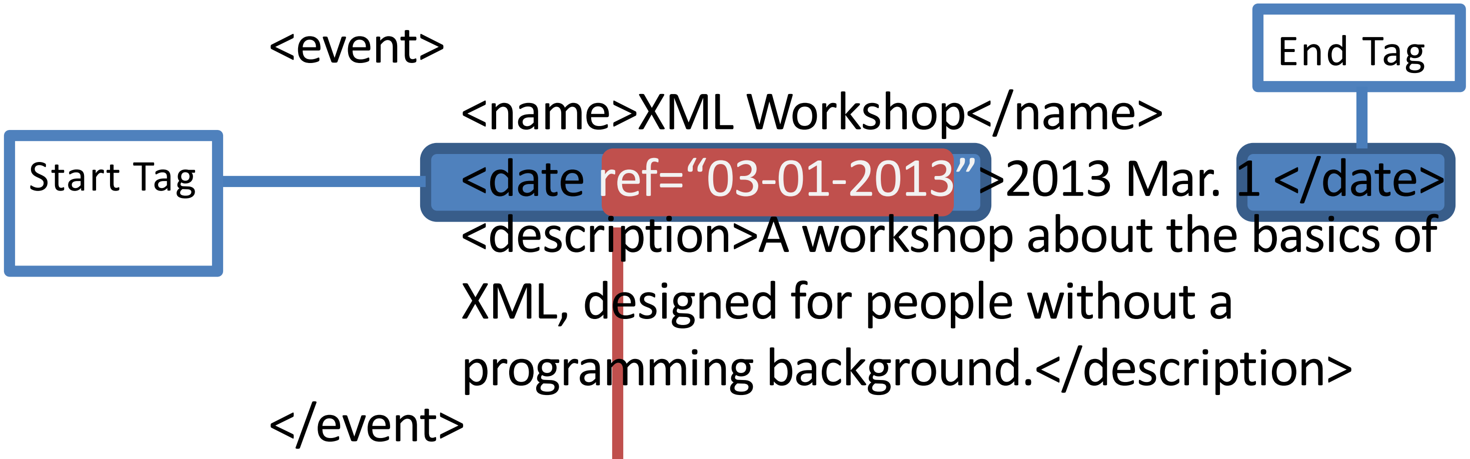


```
<event>
<name>XML Workshop</name>
<date y=“2013” m=“03” d=“01” />
<description>A workshop about the basics of XML, designed for people without a programming background.</description>
</event>
```

All of this stuff is nested (i.e. contained) within the “event” element. Elements may contain other elements and they may contain content (e.g. the prose description of the workshop).

XML Basic Concepts: Attributes and Values

Attributes and their values classify specific instances of an element.



```
<event>
      <name>XML Workshop</name>
      <date ref=“03-01-2013”>2013 Mar. 1 </date>
      <description>A workshop about the basics of
      XML, designed for people without a
      programming background.</description>
</event>
```

This is an Attribute, which has been added to the date element.

Note the syntax and the special characters: = “ ”

Also note that the attribute attaches to the Start Tag of an Element, and *not the End Tag*.

More XML Basic Concepts

Well-formed-ness

A “well-formed” XML document obeys these main rules:

1. It has at least one XML element.
2. It has a single root element, which contains all elements and content and which is not contained by any other element.
3. All elements are properly closed, with a start tag and an end tag, or as a self-closing element.
4. All elements are properly nested. If it opens inside another element, it must close inside it too.
5. All elements and attributes obey the proper syntax (e.g. no missing quotation marks).
6. No element my have the same attribute more than once.
7. (other rules) [CASE sensitive]

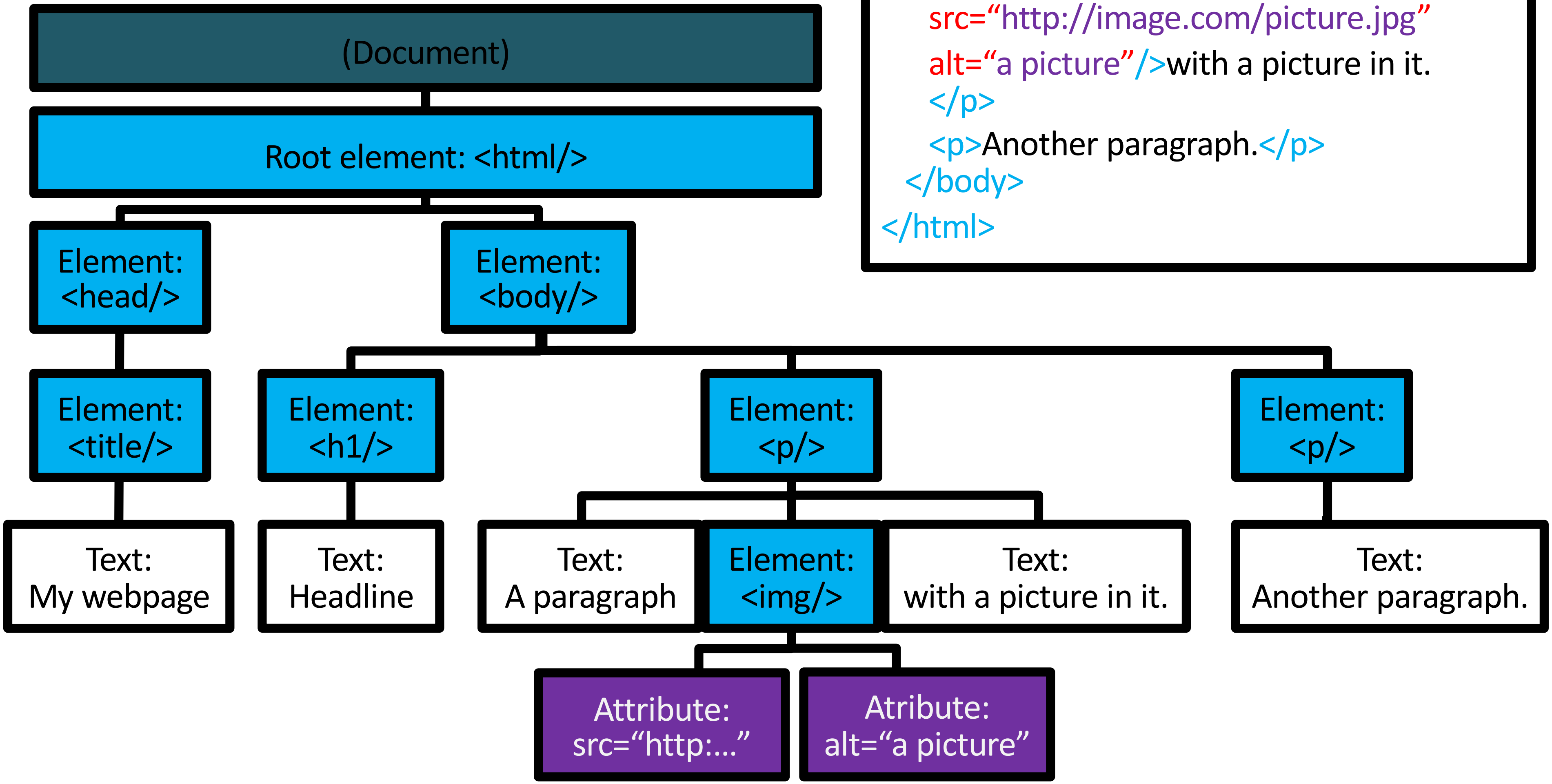
Validation

- Validation rules are additional requirements made for a particular XML application.
- Document Type Declaration (DTD) or Schema

The XML Tree

- The “Document Object Model” (DOM)
- Nodes and leaves
- Parents, Children, Siblings

```
<html>
  <head>
    <title>My Webpage</title>
  </head>
  <body>
    <h1>Headline</h1>
    <p>A paragraph<img
    src=“http://image.com/picture.jpg”
    alt=“a picture”/>with a picture in it.
    </p>
    <p>Another paragraph.</p>
  </body>
</html>
```



The XML Tree: Exercise

Working with your neighbors, draw a tree that represents the following made-up XML:

```
<book isbn=“12345”>
        <toc></toc>
        <chapter n=“1”>
                <title></title>
        </chapter>
        <chapter n=“2”>
                <title></title>
        </chapter>
        <chapter n=“3”>
                <title></title>
                <section></section>
                <section></section>
        </chapter>
        <bibl></bibl>
        <index></index>
</book>
```

XML Basics: HTML

HTML is a close sibling to XML applications, which is used for web design. It has special rules in addition to the basics of XML, for example, it must have the root element <html>, which contains all other Elements. And it must have both a <head> and a <body> element, one or the other of which contain all other elements except the root.

The Root Element <html>

<html>

<head>
</head>

The <head> Element. This contains information about the page, but will not (usually) be displayed on the page itself.

<body>

The <body> element. It contains what people see on the page.

<div>

<h1>XML Workshop
<br/>2012 Dec. 18</h1>

<p>A workshop about the basics of XML, designed for people without a programming background.</p>

</div>

</body>

</html>