XML

1

XML

- E**X**tensible **M**arkup **L**anguage
- XML was designed to ***carry data***
- XML tags are ***not predefined*** - you must define your own tags

2

XML Example

```
<message>
        <to>John</to>
        <from>Jane</from>
        <subject>Greeting</subject>
        <body>Hi John</body>
</message>
```

3

XML – Naming Rules

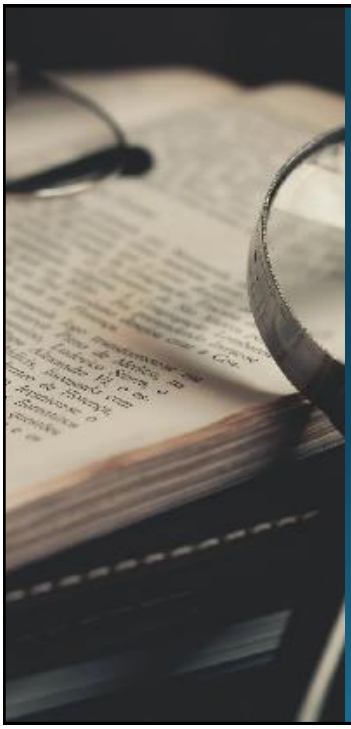
- ✓ Names can contain letters, numbers, and other characters
- ✗ Names cannot start with a number or punctuation character
- ✗ Names cannot start with the letters xml (or XML, or Xml, etc.)
- ✗ Names cannot contain spaces

4

XML Documents

- XML documents consist of three parts
  - ✓ The prolog
  - ✓ The document body
  - ✓ The epilog
- The prolog and epilog are optional and provide information about the document itself

5



Structure of an XML doc

- The prolog consists of four parts in the following order:
  - ✓ XML declaration
  - ✓ Miscellaneous statements or comments
  - ✓ Document type declaration / Schema
  - ✓ Miscellaneous statements or comments
- This order has to be followed or the parser will generate an error message.
- ***None* of these four parts are required,** but it is good form to include them.

6

XML Declaration

The XML declaration is always the **first line of code in an XML document**. It tells the processor what follows is written using XML. It can also provide any information about how the parser should interpret the code.

The complete syntax is:

```
<?xml version="version number"
      encoding="encoding type"
      standalone="yes | no" ?>
```

A sample declaration might look like this:

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


XML Comments

Comments or miscellaneous statements go after the declaration. Comments may appear anywhere after the declaration.

The syntax for comments is:

```
<!- - comment text - ->
```

This is the same syntax for HTML comments

XML Elements & Attributes

9

Elements and Attributes

Elements are the basic building blocks of XML Documents (files).

XML supports two types of **elements**:

Closed

<element_name>Content</element_name>

Example: <Student>John Doe</Student>

Empty (also called Open)

<element_name/>

Example: <Student/>

10

Elements and Attributes

Element names are ***case sensitive***

Elements can be ***nested***, as follows:

```
<message>
    <to>John</to>
    <from>Jane</from>
    <subject>Greeting</subject>
    <body>Hi John</body>
</message>
```

11

Elements and Attributes

Nested elements are called ***child elements***.

Elements must be nested correctly. Child elements must be enclosed within their parent elements.

```
<message>
    <to>John</to>
    <from>Jane</from>
    <subject>Greeting</subject>
    <body>Hi John</body>
</message>
```

All elements must be nested within a single **document** or **root element**. There can be only **one** root element.

An **open** or **empty** element is an element that contains no content. They can be used to mark sections of the document for the XML parser.

12

Elements and Attributes

An **attribute** is a feature or characteristic of an element. Attributes are text strings and must be placed in single or double quotes. The syntax is:

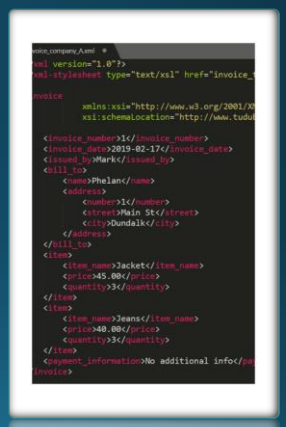
<element_name attribute="value"> … </element_name>

Example:

```
<message id="9">
    <to>John</to>
    <from>Jane</from>
    <subject>Greeting</subject>
    <body>Hi John</body>
</message>
```


Well Formed XML

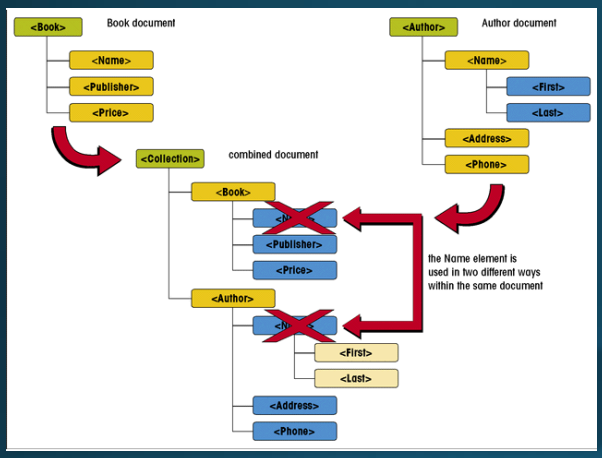
- **Well-Formed** XML is:
  - XML text that satisfies the syntactic rules as laid out in the XML specification
  - **Non well-formed** XML will cause errors in applications and parsers that expect strict XML syntax



XML Example for book data…

```
<?xml version="1.0"?>
<bookstore>
        <book category="CHILDREN">
                <title>Harry Potter</title>
                <author>J K. Rowling</author>
                <year>2005</year>
                <price>29.99</price>
        </book>
        <book category="FICTION">
                <title>The Road</title>
                <author>Cormac McCarthy</author>
                <year>2006</year>
                <price>15.99</price>
        </book>
        <book category="FICTION">
                <title>Braywatch</title>
                <author> Ross O'Carroll-Kelly </author>
                <year>2020</year>
                <price>9.99</price>
        </book>
</bookstore>
```

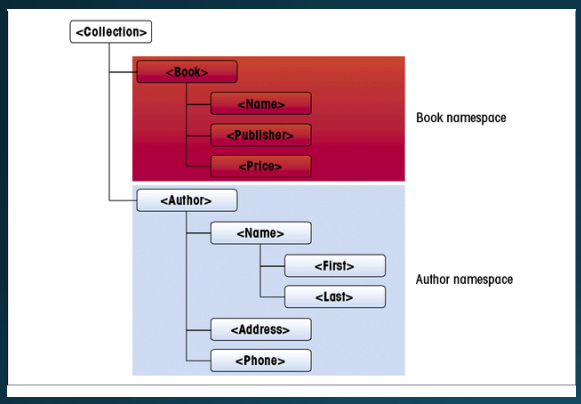

Namespaces



This figure shows two documents each with a "Name" element

Name Collision

From: New Perspectives on XML Tutorial 4



Working with Namespaces

- **Name collision** occurs when elements from two or more documents share the same name – i.e. if we combined the two XML documents to the left.
- Name collision is not a problem if you are not concerned with **validation**. The document content only needs to be **well-formed.**
- However, name collision will keep a document from being validated.
- Use **Namespaces** to avoid name Collision

Declaring a Namespace

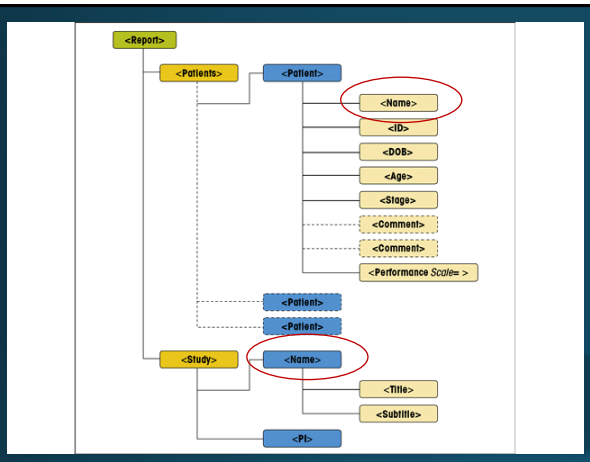
- A **namespace** is a defined collection of element and attribute names.
- Namespaces must be declared before they can be used.
- **Names that belong to the same namespace must be unique**. Elements can share the same name if they reside in different namespaces.

Declaring a Namespace

A namespace can be declared in the prolog or as an element attribute. The syntax to declare a namespace in the prolog is:

<?xml:namespace ns="*URI*" prefix="*prefix*"?>

Where *URI* is a ***U**niform **R**esource **I**dentifier* that assigns a unique name to the namespace, and *prefix* is a string of letters that associates each element or attribute in the document with the declared namespace.



This figure shows the structure of a *UHOSP.xml* file and the collision between the *Name* element.

Example: Layout of UHOSP.XML Document

From: New Perspectives on XML

Declaring a Namespace

For example,

<?xml:namespace ns="http://uhosp/patients/ns" prefix="pat"?>

- ✓ Declares a namespace with the prefix "*pat*" and the URI *http://uhosp/patients/ns*.
- ✓ The URI is not a Web address. A URI identifies a physical ***or*** an abstract resource (see next slide).

URI and URL

- A physical resource is a resource one can access and work with such as a file, a Web page, or an e-mail address. A URL is one type of URI.
- An abstract resource is one that doesn't have any physical existence, the URI is used as an identifier or an ID.

The URI *http://uhosp/patients/ns* is simply a text identifier.

Applying a Namespace to an Element

- Once it has been declared and its URI specified, the namespace is applied to elements and attributes by inserting the namespace prefix before each element name that belongs to the namespace.

```
<prefix:element>
  content
</prefix:element>
```

- Here, *prefix* is the namespace prefix and element is the **local part** of the element name.
- Prefixed names are called **qualified names** and an element name without a namespace prefix is called an **unqualified name**.

Declaring a Namespace as an Element Attribute

Qualified names can be added to a document by adding the **xmlns attribute** to an element.

The syntax is:

xmlns:*prefix*="*URI*"

Where *prefix* and *URI* are the prefix and URI for the namespace.

```
<pat:Patients xmlns:pat="http://uhosp/patients/ns">
   <Patient>
      <Name>Cynthia Dibbs</Name>
      <ID>MR890-041-02</ID>
      <DOB>1945-05-22</DOB>
      <Age>58</Age>
      <Stage>II</Stage>
      <Performance Scale="Karnofsky">0.81</Performance>
   </Patient>
</pat:Patients>
```


Declaring a Namespace as an Element Attribute (cont'd)

- The example on the previous slide applies the namespace http://uhosp/patients/ns to the Patient element *and all of its child elements*.
- While the "pat" prefix was only added to the Patients element name, the XML parser considers the other elements part of the Patients namespace and they inherit the namespace.
- They are *unqualified elements*, though, because they lack a namespace prefix.
- Declaring a namespace by adding it as an attribute of the document's root element places all elements in the namespace.
- All elements are children of the root element.

Declaring a Default Namespace

You can specify a **default namespace** by omitting the prefix in the namespace declaration.

The element containing the namespace attribute and all of its child elements are assumed to be part of the default namespace.

```
<message xmlns="http://www.1234.com">
    <to>John</to>
    <from>Jane</from>
    <subject>Greeting</subject>
    <body>Hi John</body>
</message>
```


Using Namespaces with Attributes

Attributes, like elements, can become qualified by adding the namespace prefix to the attribute name.

prefix:attribute="value"

No element may contain two attributes with the same name.

No element may contain two qualified attribute names with the same local part, pointing to identical namespaces, even if the prefixes are different.