

What is XML?

- **EXtensible Markup Language**
- A standardized method of exchanging data.
- The basis for the **Simple Object Access Protocol**
- XML is the low level formatting of data.
We should deal with data at a much higher level.

XML and HTML

- Both XML and HTML inherit from GML, the General Markup Language
- Just because they look the same doesn't mean they are the same or have the same purpose.
- XML can be embedded in HTML.
- HTML can display XML data with Cascading Style Sheets (CSS) or eXtensible Stylesheet Language (XSL)

XML format

- XML is a set of properly nesting elements.
- The elements form a tree.
- All XML files should have a header
`<?xml version="1.0" encoding="UTF-8"?>`
- The header describes the XML version and character set used.
- Header is optional.

XML Format

- There must be one and only one element that includes the entire file (*except header*)
- Elements can surround data or text.
- Elements can have attributes

```
<section level="grad">
```

```
COMP750
```

```
</section>
```

Correctness

- The available tag names are defined in a Document Type Definition (DTD) or an XML Schema.
- Programs should stop processing an XML document if there are validation errors.
- XML programs should tolerate additional embedded elements and new attributes. This makes XML files extendible.

XML Schemas define

- elements that can appear in a document
- attributes that can appear in a document
- which elements are child elements
- the order of child elements
- the number of child elements
- whether an element is empty or can include text
- data types for elements and attributes
- default and fixed values for elements and attributes

Namespaces

- To avoid tag name conflicts, you can prefix a tag name with a namespace name.
- The namespace attribute can be included in any element tag
xmlns:prefixname="namespace"
- The **namespace** is a URL, but it has no programmatic connection to the namespace
- Namespaces apply within the scope of the element containing the `xmlns` attribute

Namespace Example

<myroot>

<Z xmlns:a="http://williams.comp.ncat.edu/comp750">

<a:X >This is the data in the element </a:X>

<a:Y>more data with <bracket</a:Y>

</Z>

</myroot>

Special Characters

- There are 5 predefined entity references in XML:
 - < < less than
 - > > greater than
 - & & ampersand
 - ' ‘ apostrophe
 - " “ quotation mark
- `<![CDATA[" stuff "]]>` allows you to include data that is not parsed.

Microsoft XML Notepad

- Microsoft utility to view and create XML files.
- available for free from
[http://www.devhood.com/tools/
tool_details.aspx?tool_id=261](http://www.devhood.com/tools/tool_details.aspx?tool_id=261)

Challenges with XML

- Size
 - Converting a simple database into XML can increase its size by four fold or more.
- Native Use
 - XML is probably not a suitable format for storing a database. Data received in XML will probably have to be converted.
- Immature standard
 - XML is still changing.