Meta Information

BSc Information Technology – Yr 2

Semester 2

Dr. Brenda Mullally

Syllabus

Introduction

Web 2.0

Web information management

XML

Extensible style sheets (XSLT)

Introduction

Anywhere there is information, you'll find XML.

What is XML?

What can I do with XML?

How do I get started?

Introduction

What is XML? – An Information Container

XML contains, shapes, labels, structures, and protects information.

XML is HTML plus

Tag it, display it.

Also treat it like

A word processing file – type it, edit it, correct it.

A database – open it, search it, add, update, delete.

Introduction

XML contains information using symbols called markup.

Markup enhances the meaning of information, identifying parts and how they relate to each other.

Newspaper articles.

Markup is important to electronic documents because they are processed by computer programs. If a document has no labels or boundaries, then a program will not know how to distinguish a piece of text from any other piece.

XML's markup divides a document into separate information containers called elements.

Introduction

These "elements" contain the information, they seal it up, label it, and provide a convenient package for computer processing.

Like boxes, elements nest inside other elements.

One large element may contain a whole lot of other elements, which in turn may contain other elements and so on down to the data.

This creates an unambiguous hierarchical structure that preserves information such as sequence, ownership, position, description, association.

Introduction

```xml
<?xml version="1.0"?>

<telegram pri="important">
   <to>Sarah Bellum</to>
   <from>Colonel Timeslip</from>
   <subject>Robot-sitting instructions</subject>
   <graphic fileref="figs/me.jpg"/>
   <message>Thanks for watching my robot pal
     <name>Zonky</name> while I'm away.
     He needs to be recharged <emphasis>twice a
     day</emphasis> and if he starts to get cranky,
     give him a quart of oil. I'll be back soon,
     after I've tracked down that evil
     mastermind <villain>Dr. Indigo Riceway</villain>.
   </message>
</telegram>
```

Can you tell the difference between the markup and the data?

Introduction

The markup symbols are delineated by angle brackets <>.

<to> and </villian> are examples of tags.

The data, or content, fills the space between these tags.

As you get used to xml you will use the tags as signposts for finding your way through a document.

At the top of the document is the XML declaration,

<?xml version="1.0"?>

This helps the XML processing program identify the version and what kind of character encoding it has. It is optional but a good thing to include.

Introduction

Next, we see the <telegram> tag. This is the start of an element. We say that the element's name or type is "telegram", or you can call it a "telegram element". The end of the telegram element is at the bottom of the document shown by the </telegram> tag.

This first element contains all of the contents of the document. Thus we call it the document element (or the root element).

Inside you see more element with start tags and end tags following a similar pattern.

Introduction

There is one exception, the empty tag

<graphic fileref="figs/me.eps"/>

This represents an empty element, rather than containing data, this element references some other information that should be used in its place, in this case a graphic is to be displayed.

Every element that contains data must have a start tag and an end tag. The name of the start and end tag must match exactly. XML is case sensitive and very picky about details.

Introduction

Some tags function as bookends, marking the beginning and ending of regions, while others mark a place in the text.

Boundaries: <telegram> and </telegram> define the start and end of a collection of text and markup.

Roles: What is a region of text doing in the document? <name> indicates a purpose for the content of the element.

Positions: Elements preserve the order of their contents.

Containment: Nesting is taken into account, how a <name> element is treated in a <message> may be different if it was in a <to> element.

Relationships: The <graphic../> tag forms a link to a file.

Introduction

In XML a document is a general term, it is the basic unit of XML information, composed of elements and other markup in an orderly package.

It can contain text such as a story or article but it doesn't have to. It might consist of a database of numbers, or some abstract structure representing a molecule or equation.

A document is also not necessarily a file. A file is the physical structure. An XML document can exist in one file or in many files, some of which may be on another system. It may not be a file at all but generated from a program. An XML document is a logical structure.

12

Introduction

What's involved?

The instance (.XML)

     Your database of information

The schema (.XSD)

     The rules behind the database

The IDE (Integrated design environment)
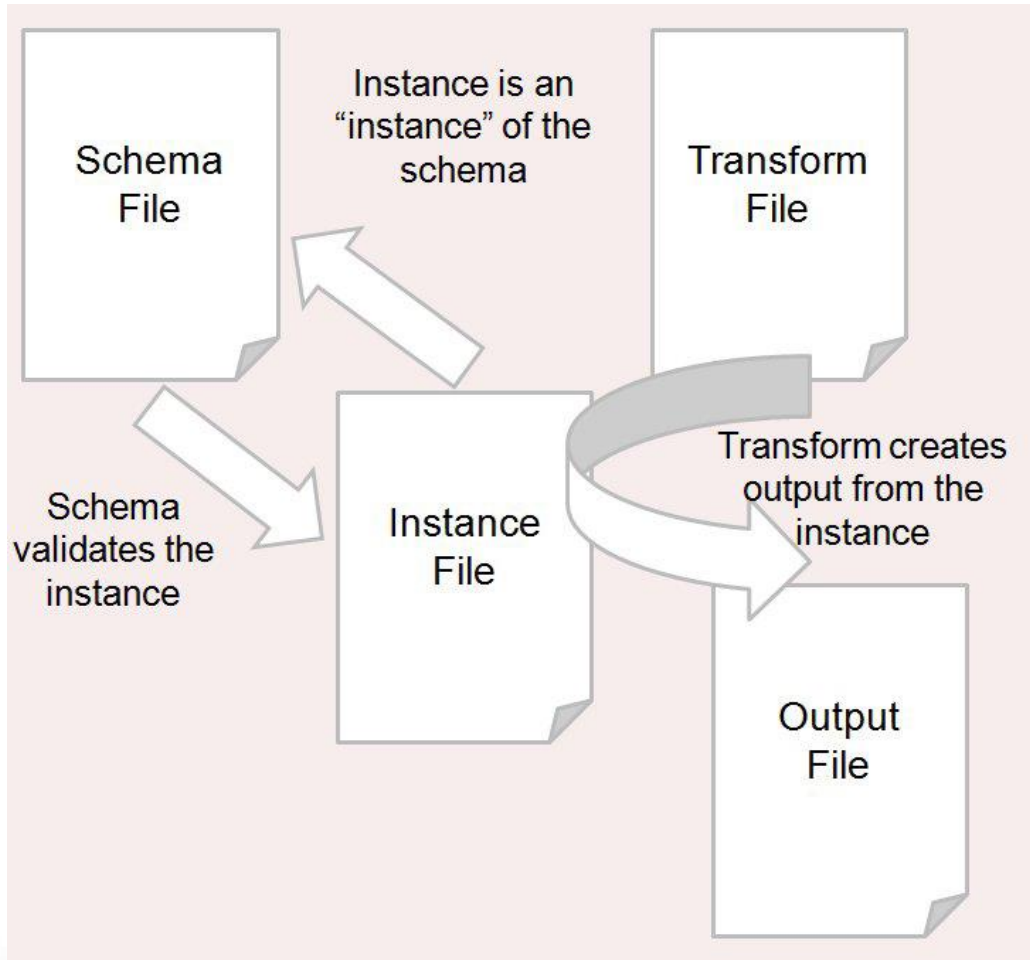
     To create, edit, validate, transform

The transformation (.XSL)

     Selects the right information from the instance and presents   it to the user.

The output (.HTML, .PDF, etc)

     Presents one "cut" of the information to the user.

Introduction

Schema File

Instance is an "instance" of the schema

Transform File

Schema validates the instance

Instance File

Transform creates output from the instance

Output File

An instance contains your base information, it is guided by strict rules, and can be edited.

Introduction

Goals of XML:

Form should follow function, XML document created for the purpose specific to it's data. Rather than creating a language to contain all types of data, users choose element names and decide arrangement in the document.

Unambiguous, A document should be marked up so there is only one way to interpret the names, order, and hierarchy of elements. XML rules about syntax are strict.

15

Introduction

Goals of XML:

Separate markup from presentation, style should always be separate from document.

<example> Goethe once said, <i>Lieben ist wie</i>

</example>

<example> Goethe once said, <foreignphrase>Lieben is wie</foreignphrase>

Introduction

Goals of XML:

Keep It Simple, easy to read, easy to process

Maximum error checking, well-formed conforms to rules.

Inclusive, Unicode is used for the character set, opening it up to thousands of letters, ideographs, and symbols.

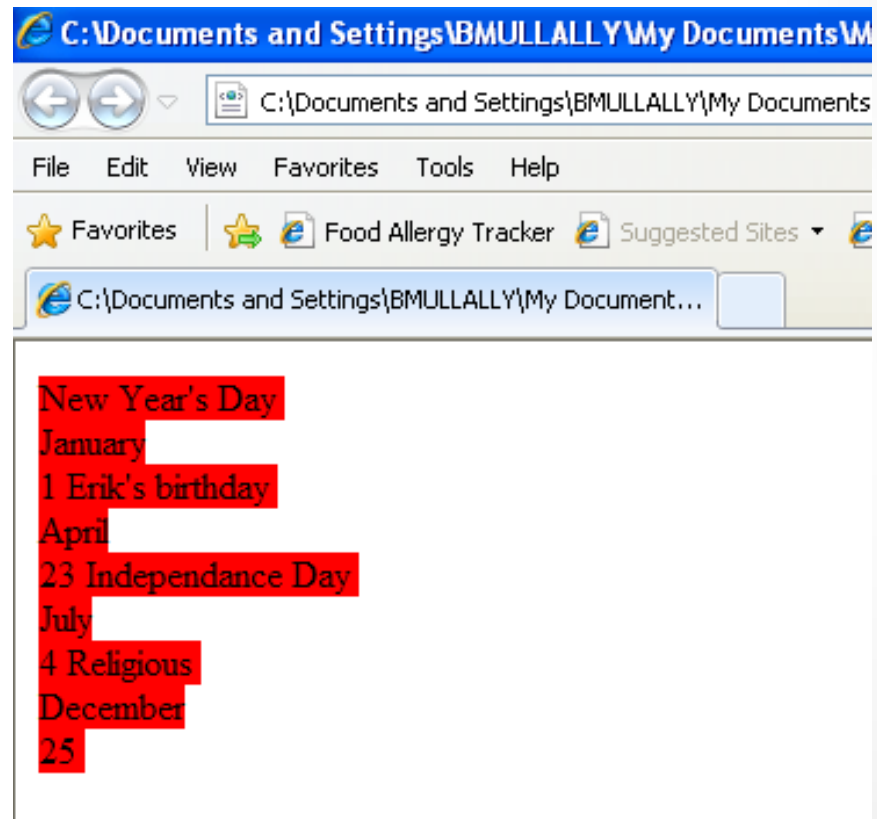# Introduction

What can I do with XML?

Store and retrieve data: Just about every software application needs to store some data. Hard coding data in web pages or any program is not maintainable. Store data separately and your program simply reads the xml document and processes it. XML is appropriate for data that is not searched randomly, it is primarily a sequential storage medium, a search moves through the document. It is appropriate for a novel but not a phone book.

Introduction

Format Documents: XML documents are processed to put the data in a presentable form.

CSS – Simplest is to apply a css file to it. This is only used where you are presenting the data in the order it is stored. It is not used where you only want to present an excerpt of the data. Calendar.xml calendar.css

Introduction

Transformation- is a process of breaking apart the XML document and building a new one. The new document may or may not use the same markup language. You can sort elements, remove those you don't want, generate new data such as headers or footers. This is typically done using XSLT, essentially a programming language optimised for transforming XML. It requires a transformation instruction which happens to be called a stylesheet (not same as css).

Introduction

Transforming XML into HTML is fine for online viewing. IT is not so good for print media. HTML was never designed for handling the complex formatting of printed documents, with headers and footers, multiple columns, page breaks. For that you need to transform into a richer format such as PDF.

Direct transformation to PDF is not easy to do. It requires extensive knowledge of the PDF specification which is huge and difficult.

Introduction

There are many presentational formats, Adobe's PostScript and PDF and Microsoft's Rich Text Format (RTF) are all used.

Programming solutions:

If style sheets do not fit the bill, then you may find a programming solution is necessary.

Although XSLT has much to offer in transformation, it tends to be rather weak in some areas, such as processing character data. It is sometimes necessary to pre-process XML data before a transformation, or to write a program that does the whole processing form source to presentational format.

XML Advantages

- Simplicity
- Open standard and platform/vendor-independent
- Extensibility
- Reuse
- Separation of content and presentation
- Improved load balancing
- Support for integration of data from multiple sources
- Ability to describe data from a wide variety of applications
- More advanced search engines
- New opportunities.

XML Advantages

Simplicity:
XML is a relatively simple standard, less than 50 pages long. It was designed as a text based language that is human-legible and reasonably clear.

Open standard and platform/vendor independent:
XML is both platform independent and vendor independent, a restricted form of SGML and an ISO standard. It is based on ISO 10646, the Unicode character set, so it has built in support for texts in all the world's alphabets.

XML Advantages

Extensibility:
Unlike HTML, XML is extensible allowing users to define their own tags to meet their own particular application requirements.
Reuse:
Extensibility allows libraries of XML tags to be built once and reused by many applications.

Separation of content and presentation:
XML separates the content of a document from how the document will be presented. This facilitates a customised view of the data, perhaps based on factors such as the user preference or configuration.

XML Advantages

Improved load balancing:
Data can be delivered to the browser on the desktop for local computation, thus moving computation from the server and achieving better load balancing.

Support for the integration of data from multiple sources:
XML offers the ability to integrate data from many sources.As XML is extensible, it can be used to describe data contained in a wide variety of applications. XML makes the data self-describing, the data can be received and processed without the need for a built-in description of the data.

XML Advantages

More advanced search engines:

With XML, search engines will be able to parse the description bearing tags.

New opportunities:

Perhaps the great advantage of XML is the wealth of opportunities that are now presented by this technology.

Introduction

World Wide Web Consortium W3C
W3C Founded in 1994 by Tim Berners-Lee   www.w3.org

The organisation was devoted to developing nonproprietary, interoperable technologies for the World Wide Web thus making the Internet universally accessible regardless of ability, language or culture.
It is also a standardisation organisation.

Technologies standardised by W3C include:
- Cascading Style Sheets (CSS).
- HyperText Markup Language (HTML).
- Extensible Markup Language (XML).

Introduction

Resources:

http://www.w3.org/standards/xml/

http://www.deitel.com/ResourceCenters/Programming/XML/tabid/279/Default.aspx

http://xml.silmaril.ie/whatfor.html

http://www.w3schools.com/XML/