XML

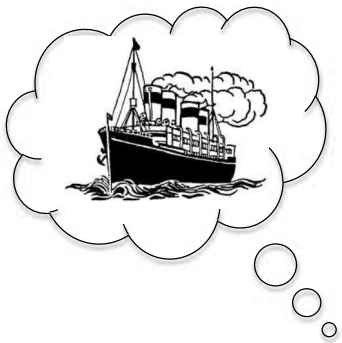
XML - I

Module 1

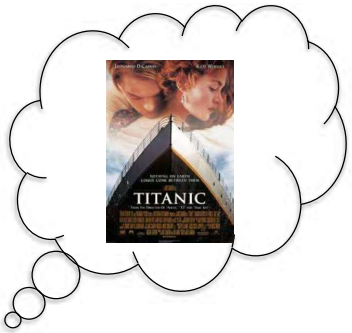
Introduction to XML

How to understand data

- What comes to your mind you hear the word ‘TITANIC’?



Will you think about the Ship or the Movie???



How to understand data

- Here we are presented with a problem:

When you specify something, like ‘Titanic’, there should be a way to find out whether you meant the ship or the movie.

How to describe data?

This is what we are going to learn here…

What is a Document

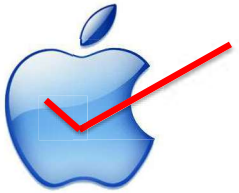
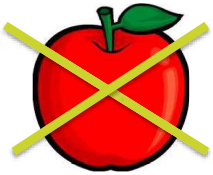
- Before we start learning XML, we will have a look at the need to have such a language. We begin with understanding what is a document.

A document is a combination of data and markup.

- What is data and what is markup?
- Imagine that you are reading a e-book. Find all the occurrences of the word 'Apple' in the book.
- Using 'find' option, we can easily find the word wherever it is present in the document.

What is a Document

- But now what if we need to find all the occurrences of the word 'Apple' where it is meant as the 'Apple' company and not the fruit.


In a plain text document, finding a word with a specific meaning is difficult!!!
Especially if that word has more than one meaning!
The same problem occurs during data processing also…

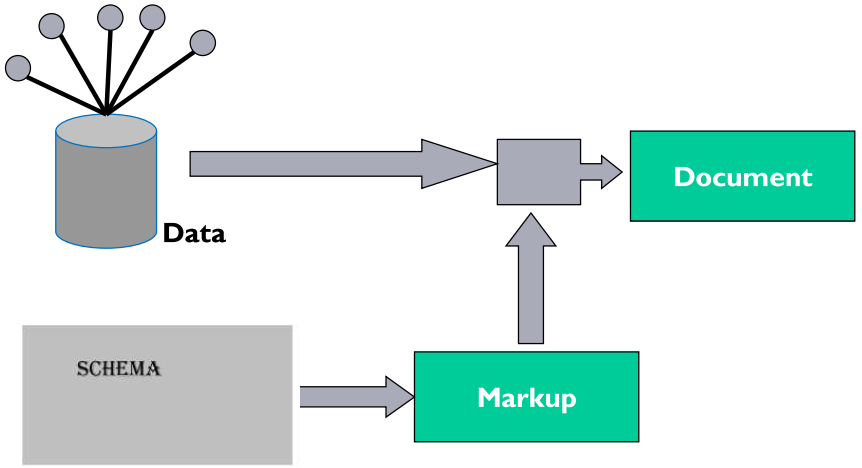
Role of Markup

- **Markup describes the data in the document and how it should be interpreted.**

Markup is anything inside a pair of angle brackets (< >)

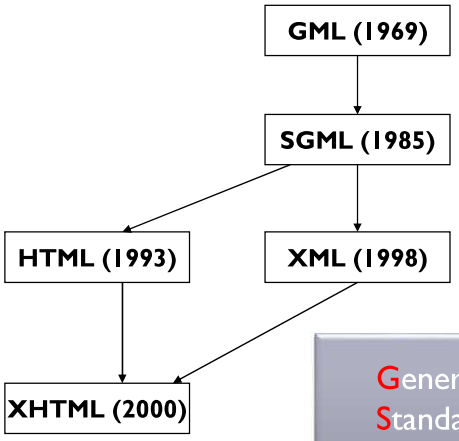
- **Need for markups**
  - **The basic purpose of going for a markup is to guarantee " What You See Is What You Get" Format. Meaning, there should not be any discrepancies in display even across dissimilar systems at any given time.**
  - **Data Transfer across various dissimilar applications**
- **A markup language must specify**
  - What markup is allowed
  - What markup is required
  - How markup is to be distinguished from text

What is a Document



A document is a combination of data and markup. It is very similar to a person reading a book. While reading one markup's information he finds it relevant. Similarly, its an effort to mark relevant data. The data could be pulled out of any relational schema or any application.

Markup Languages



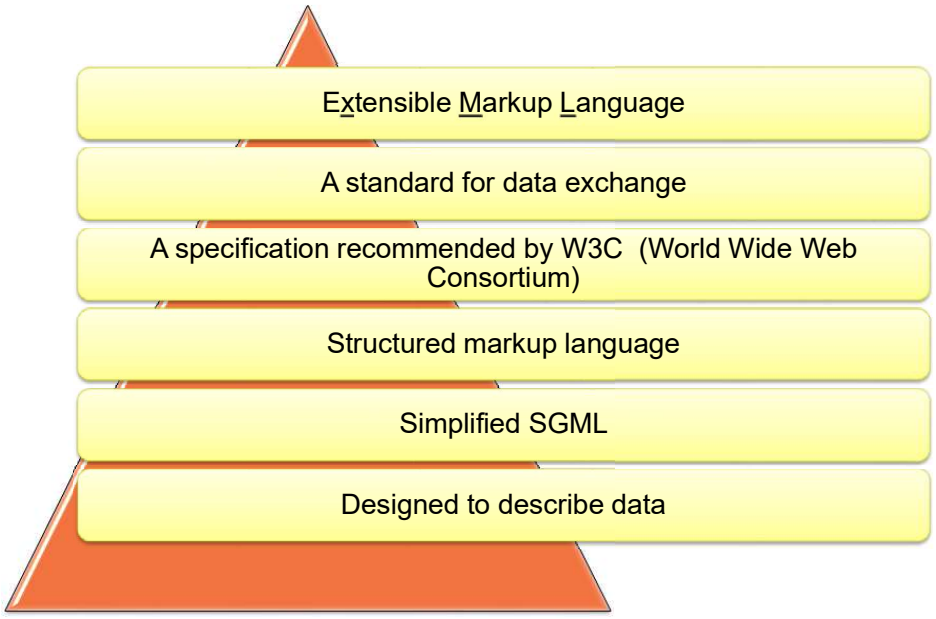
The concept of markup languages was initially implemented by IBM in 1969 with the development of the Generalized Markup Language by Goldfrab, Mosher, Lorie.

Generalized Markup Language
Standard Generalized Markup Language
HyperText Markup Language
eXtensible Markup Language
eXtensible HyperText Markup Language

HTML – Hyper Text Markup Language

- Primarily designed for desired look and feel
- Intended to remove difficulties of GML and SGML
- A non structured markup language
- Did not support what is called as 'Content Markup'.
- Both presentation markups and structural markups were placed in the same Document.

What is XML?



HTML and XML – A comparison

HTML	XML
Used to mark up text so that it can be displayed to the user	Used to mark up data so that it can be processed by applications/computers
Uses fixed set of predefined tags	You can create your own tags in XML
Describes appearance as well as structure of the data	Describes the content
Is for humans. So, most browsers correct or ignore as many errors as they can	Is for computers. So, the rules are strict and errors are not allowed
Designed to display data and to focus on how data looks	Designed to describe data and to focus on what data is
HTML describes both structure (e.g. <p>, <h2>) and appearance (e.g. , <font>, <i>)	

Uses of XML

Can separate data from HTML

With XML, your data is stored outside your HTML

Used to exchange data

With XML, data can be exchanged between incompatible systems

Can be used to share data

With XML, plain text files can be used to share data

Can be used to store data

With XML, plain text files can be used to store data.

Example of an XML Document

```
<?xml version="1.0"?>
<Person>
<Name>Jessy</Name>
<Age>12</Age>
<Country>India</Country>
</Person>
```

Prologue

Root Element

Document

XML Building Blocks - Prolog

- A part of an XML document that precedes XML data
- It includes
  - A declaration: version [, encoding, standalone]
  - An optional DTD (Document Type Definition)*
- Example of an XML declaration

```
<?xml version="1.0" encoding="ISO-8859-1"
standalone="yes"?>
```

* Note: DTD will be discussed later

XML Building Blocks

The XML document is also built from:

- Elements: Basic unit of an XML document
  - An element is a logical structure in XML that is delimited by a start tag and an end tag
  - Element consists of 3 parts: Start Tag, Content and End Tag

<NAME> JESSY </NAME>

Start Tag | Content | End Tag

Rules:
- Start Tag and End Tag should match and it is case sensitive
- Content cannot contain < or &

XML Building Blocks

- Attributes: provides additional information
  - Specified in the start tag of the element
  - Have a key- value pair
  - An element can have any number of attributes
  - Attribute can have only one value

<FLOWER COLOR="RED" > ROSE </FLOWER>

Attribute Name

Attribute Value

Rules:
- Attribute value should be enclosed in double quotes
- Attribute names within an element are unique
- attribute names must begin with a letter or underscore and can contain letters, digits, underscore(_), dot(.), hyphen(-)

Well-formed XML document

An XML document is said to be well-formed if it follows basic syntax rules specified for XML by W3C

XML tags are case sensitive

Must have only one root element

Every element must have a closing tag

Elements must be properly nested

Attribute values must always be quoted (single or double)

Well-formed XML document

- Others:

Comments in XML: <!-- This is a comment -->

Every element must have a start tag and an end tag. However empty elements can end with /> instead of > (e.g. : <book id ="001" />)

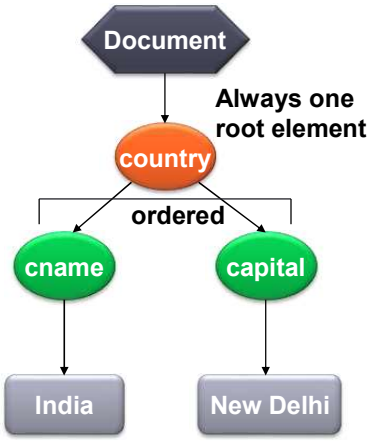
Colon(:) is used for namespaces.

XML Tree Structure

XML Document

```
<country>
    <cname>India</cname>
    <capital>New Delhi</capital>
</country>
```

Document tree



Assignment

Summary

In this module, you were able to

- Explain evolution of XML
- Describe XML Structure
- Understand well-formedness of documents

WIPRO
Applying Thought

Thank You