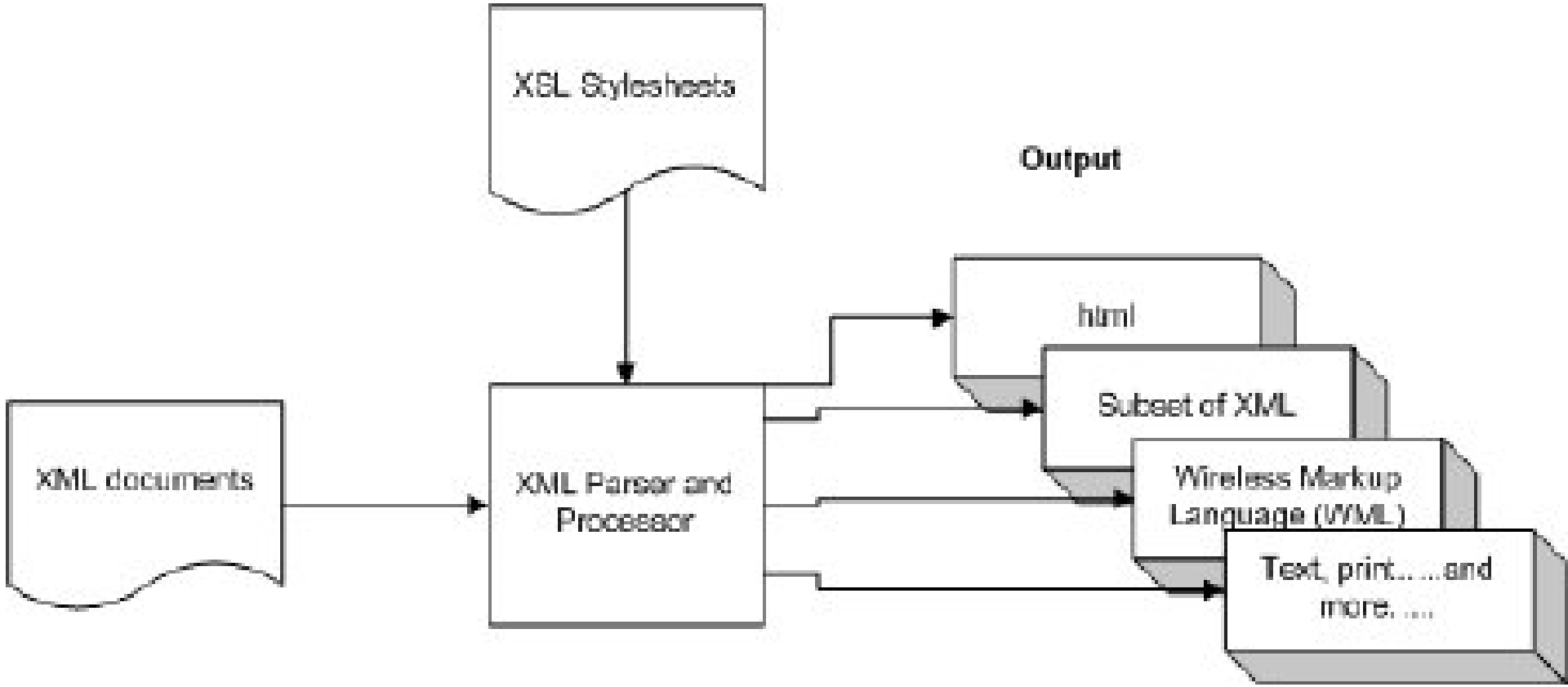
XML Technologies

Internal : 90 Marks

Moodle Test : 30 (3 Moodle Test)
Written Exam : 20
Practical : 20
Viva : 10
Assignment / Class Work : 10

External: 60 Marks

The basic XML flow



What is XML?

XML stands for eXtensible Markup Language.

Xml is a mark up language.(A **mark up language** is a modern system for highlight or underline a document.)

XML is designed to store and transport data.

Xml was released in late 90's. it was created to provide an easy to use and store self describing data.

XML became a W3C Recommendation on February 10, 1998.

XML is not a replacement for HTML.

XML is designed to be self-descriptive.

XML is designed to carry data, not to display data.

XML tags are not predefined. You must define your own tags.

XML is platform independent and language independent.

Note: Self-describing data is the data that describes both its content and structure.

What is XML?
a family of technologies

No	Technology	Meaning	Description
1)	XHTML	Extensible html	It is a clearer and stricter version of XML. It belongs to the family of XML markup languages. It was developed to make html more extensible and increase inter-operability with other data.
2)	XML DOM	XML document object model	It is a standard document model that is used to access and manipulate XML. It defines the XML file in tree structure.
3)	XSL it contain three parts: i) XSLT (xsl transform) ii) XSL iii)XPath	Extensible style sheet language	i) It transforms XML into other formats, like html. ii) It is used for formatting XML to screen, paper etc. iii) It is a language to navigate XML documents.
4)	XQuery	XML query language	It is a XML based language which is used to query XML based data.
5)	DTD	Document type definition	It is an standard which is used to define the legal elements in an XML document.
6)	XSD	XML schema definition	It is an XML based alternative to dtd. It is used to describe the structure of an XML document.
7)	XLink	XML linking language	Xlink stands for XML linking language. This is a language for creating hyperlinks (external and internal links) in XML documents.
8)	XPointer	XML pointer language	It is a system for addressing components of XML based internet media. It allows the xlink hyperlinks to point to more specific parts in the XML document.
9)	SOAP	Simple object access protocol	It is an acronym stands simple object access protocol. It is XML based protocol to let applications exchange information over http. in simple words you can say that it is protocol used for accessing web services.
10)	WSDL	web services description languages	It is an XML based language to describe web services. It also describes the functionality offered by a web service.
11)	RDF	Resource description framework	RDF is an XML based language to describe web resources. It is a standard model for data interchange on the web. It is used to describe the title, author, content and copyright information of a web page.
12)	SVG	Scalable vector graphics	It is an XML based vector image format for two-dimensional images. It defines graphics in XML format. It also supports animation.
13)	RSS	Really simple syndication	RSS is a XML-based format to handle web content syndication. It is used for fast browsing for news and updates. It is generally used for news like sites.

XML Facts

- officially recommended by W3C since 1998
- a simplified form of SGML (Standard Generalized Markup Language)
- primarily created by Jon Bosak of Sun Microsystems
- important because it removes two constraints which were holding back Web developments:
  1. dependence on a single, inflexible document type (HTML);
  2. the complexity of full SGML, whose syntax allows many powerful but hard-to-program options

HTML and XML

- HTML

-HTML is used to mark up text so it can be displayed to users

-HTML describes both structure (e.g. <p>, <h2>,) and appearance (e.g.
, <font>, <i>

-HTML uses a fixed, unchangeable set of tags

- XML

-XML is used to mark up data so it can be processed by computers

-XML describes only content, or "meaning"

- allows user to specify what each tag and attribute means
- In XML, you make up your own tags

HTML and XML

- HTML and XML look similar, because they are both SGML languages (SGML = Standard Generalized Markup Language)
  - Both HTML and XML use elements enclosed in tags (e.g. <body>This is an element</body>)
  - Both use tag attributes (e.g., <font face="Verdana" size="+1" color="red">)
  - Both use entities (<, >, &, ", ')
- More precisely,
  - HTML is defined in SGML
  - XML is a (very small) subset of SGML

HTML and XML

- HTML is for humans
  - HTML describes web pages
  - You don't want to see error messages about the web pages you visit
  - Browsers ignore and/or correct as many HTML errors as they can, so HTML is often sloppy

- XML is for computers
  - XML describes data
  - The rules are strict and errors are not allowed
    In this way, XML is like a programming language
  - Current versions of most browsers can display XML
    However, browser support of XML is spotty at best

Problems with HTML -

- No syntax checking
  - No provision for validating HTML documents
- No structure
  - Display-related characteristics are considered and nothing else
- Not content-aware
  - Use of tags such as <H3> instead of <Name>
- Not international
  - Based on ASCII, so limited language support (XML is Unicode-based)
- Not suitable for data interchange
  - Very little information inside a document is identified
- Not object-oriented
  - Programmers with OO skills find it difficult to deal with HTML

What is XML?

- *eXtensible* - By applying Identifiers for elements of information in a neutral way, stored in a neutral form, independent of systems, devices and applications.

- *Markup* - For adding information to a document relating to its structure and/or content.

What is XML?

- *Language* - A standard methodology with formal syntax……..However
- XML is not a language but a toolkit for developing or defining languages
- A common syntax for expressing structure in data

XML is Tags and Content

A Hierarchical Data Representation

A Tree Structure Data Representation

Example:

Tags

```
<class>
<student>
<name>ABC</name>
<GPA>3.5</GPA>
</student>
</class>
```

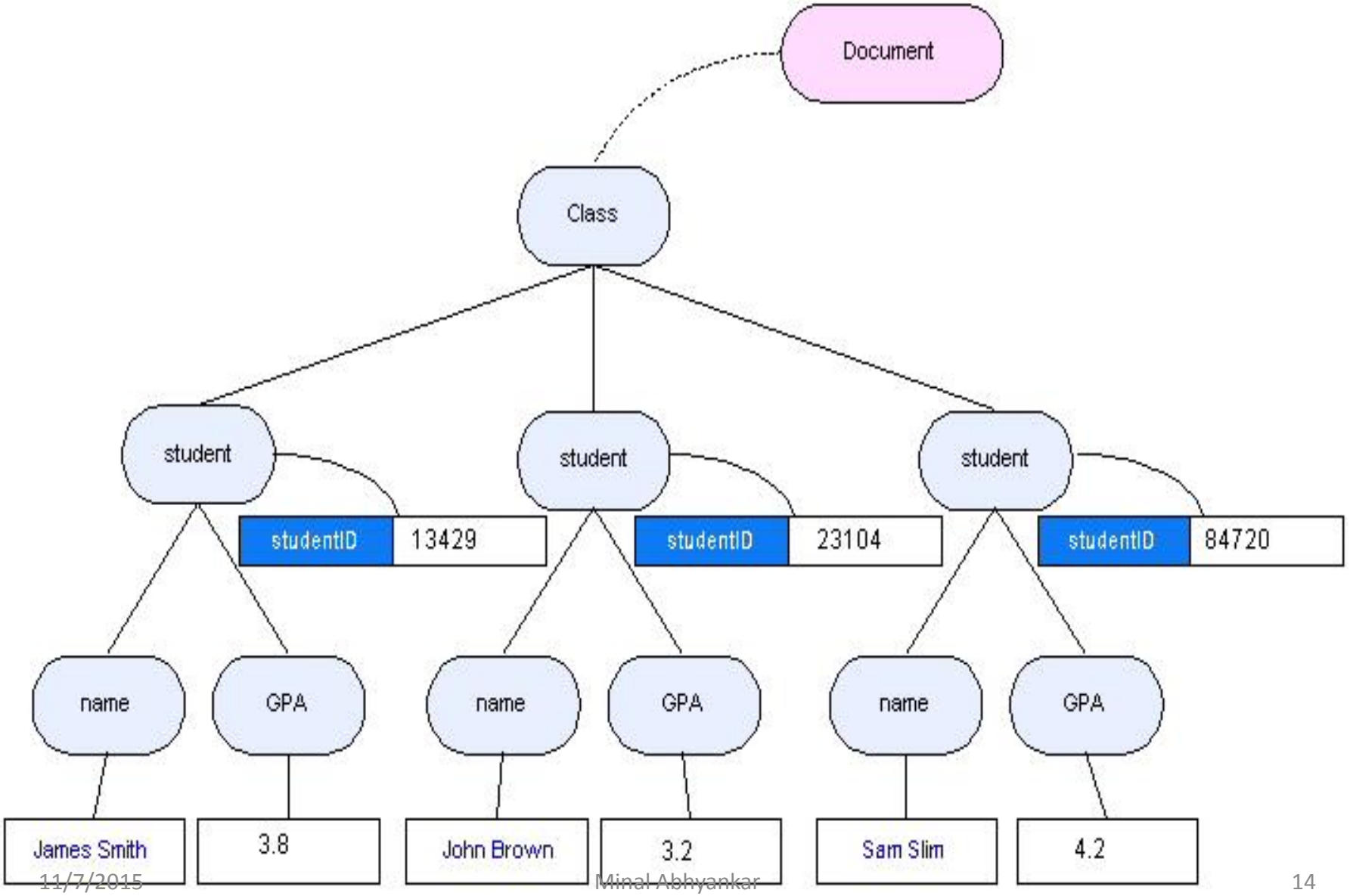
Content

```
<class>
<student>
<name>ABC</name>
<GPA>3.5</GPA>
</student>
</class>
```

Example of XML :

```
<?xml version="1.0" encoding="UTF-8"?>
<!-- This is a comment -->
<class>
    <student studentID='13429'>
        <name>James Smith</name>
        <GPA>3.8</GPA>
    </student>
    <student studentID='23104'>
        <name>John Brown</name>
        <GPA>3.2</GPA>
    </student>
    <student studentID='84720'>
        <name>Sam Slim</name>
        <GPA>4.2</GPA>
    </student>
</class>
```

Tree structure of the XML sample



Why Do We Need XML?

- Interoperability
- Info can be separated from the way it is presented
- Information can be presented in a variety of ways depending on the device the end user has.

Why Is XML Important?

- Plain Text
  - Easy to edit
  - Useful for storing small amounts of data
  - Possible to efficiently store large amounts of XML data through an XML front end to a database
- Data Identification
  - Tell you what kind of data you have
  - Can be used in different ways by different applications
- Stylability
  - Inherently style-free
  - XSL---Extensible Stylesheet Language
  - Different XSL formats can then be used to display the same data in different ways
- Inline Reusabilty
  - Can be composed from separate entities
  - Modularize your documents without resorting to links

Why Is XML Important?

- Linkability -- XLink and XPointer
  - Simple unidirectional hyperlinks
  - Two-way links
  - Multiple-target links
  - “Expanding” links
- Easily Processed
  - Regular and consistent notation
  - Vendor-neutral standard
- Hierarchical
  - Faster to access
  - Easier to rearrange

The Basic Rules –

- XML is case sensitive
- All start tags must have end tags
- Elements must be properly nested
- XML declaration is the first statement
- Every document must contain a root element
- Attribute values must have quotation marks
- Certain characters are reserved for parsing

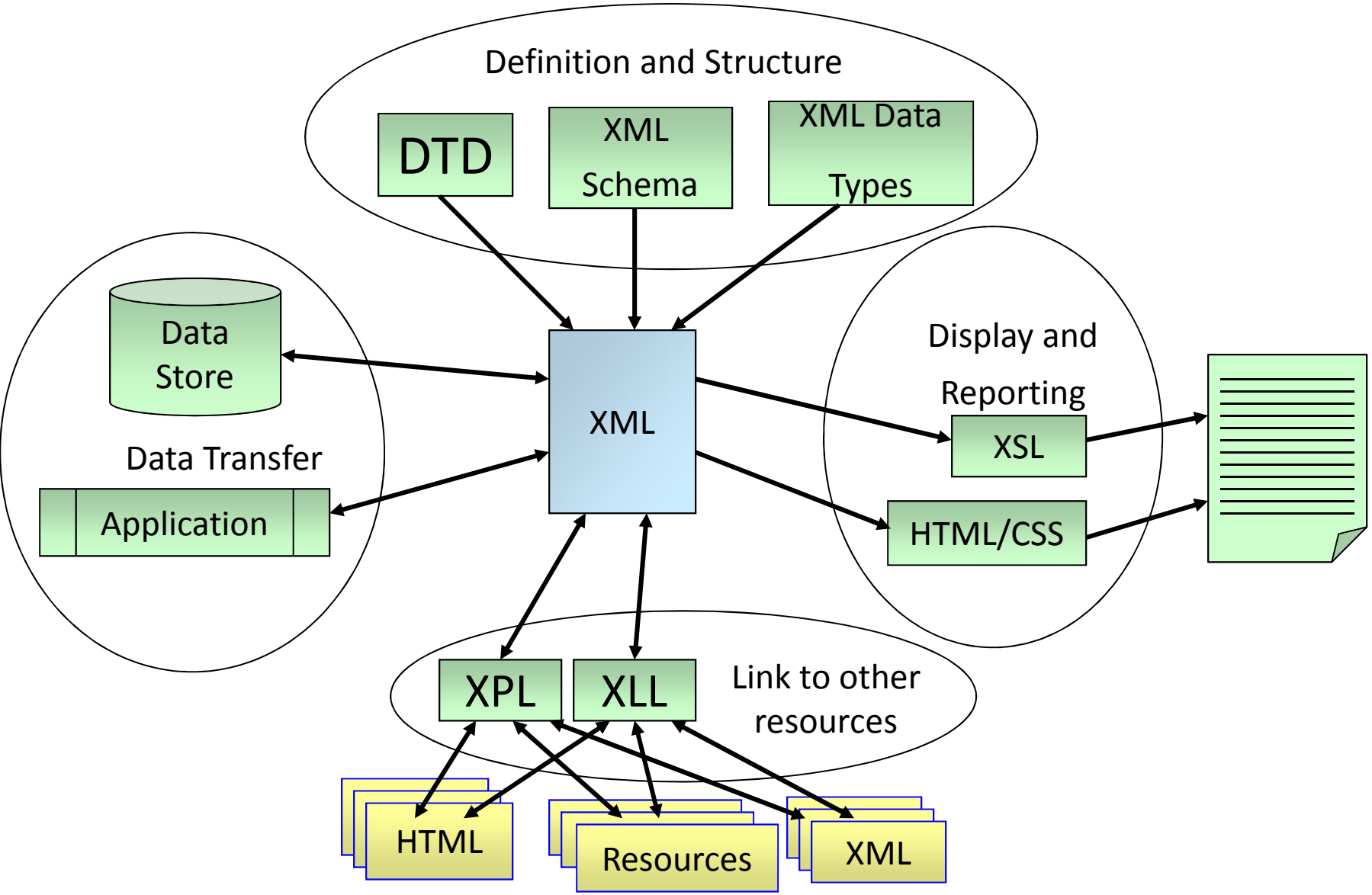
How Does It Work?

- XML Parser
- XML Processor
- XML does not provide an application programming interface (API). (Just passes data to the application)

XML Related Technologies

- DTD (Document Type Definition) and XML Schemas are used to define legal XML tags and their attributes for particular purposes
- CSS (Cascading Style Sheets) describe how to display HTML or XML in a browser
- XSLT (eXtensible Stylesheet Language Transformations) and XPath are used to translate from one form of XML to another
- DOM (Document Object Model), SAX (Simple API for XML, and JAXP (Java API for XML Processing) are all APIs for XML parsing

XML Technologies



Viewing XML

- XML is designed to be processed by computer programs, not to be displayed to humans

Nevertheless, almost all current browsers can display XML documents

- They don't all display it the same way
- They may not display it at all if it has errors
- For best results, update your browsers to the newest available versions

- Remember:
  - HTML is designed to be *viewed,*
  - XML is designed to be *used*

Extended document standards

You can define your own XML tag sets, but here are some already available:

- XHTML: HTML redefined in XML
- SMIL: Synchronized Multimedia Integration Language
- MathML: Mathematical Markup Language
- SVG: Scalable Vector Graphics
- DrawML: Drawing MetaLanguage
- ICE: Information and Content Exchange
- ebXML: Electronic Business with XML
- cxml: Commerce XML
- CBL: Common Business Library

XML Document Contents

An XML document is composed of

1. Declarations (prolog, dtd Reference)
2. Comments
3. Elements

Markup

4. Entities (predefined ,custom defined ,character entities)

Content

Processing Instruction

- A type of tag supported by XML
- Declares information necessary for processing a document, or directs any program that processes the document to perform a specific function
- Starts with <? and ends with ?>
- Example
  - <?xml version="1.0"?>
- Provides information about a specific application

- The <?xml version="1.0" ?> declaration is necessary for an software to identify an xml document.
- The other attributes of this processing instruction are :
  - encoding : An optional attribute that specifies the character encoding schema used in this XML file. Currently, there are many encodings supported by most XML applications: UTF-8, UTF-16, ISO-10646-UCS-2, ISO-10646-UCS-4, ISO-8859-1, ..., ISO-8859-9, ISO-2022-JP, Shift_JIS, EUC-JP. Default is UTF-8.
  - standalone : specifies if this xml doc also refers to other external docs like dtd or css/xsl etc or is self sufficient. If standalone is yes, it cannot refer to any other dtd or stylesheet file.
  - Version is an mandatory attribute.

Comments

Ex: <!-- This is a comment -->

- Double hyphen '--' must not occur within comments.
- Comments cannot be nested.
- You can use a comment anywhere in XML document except within attribute value.
- Comment declaration should not appear before XML.

Root Tag

- After declarations, every XML file has exactly one element, known as root element. Any other elements in file are contained within that element.

```
<?xml version='1.0' encoding='utf-8'?>
<!-- A SAMPLE set of slides -->
<slideshow>   </slideshow>
```

Elements in XML

- An XML element is everything from (including) the element's start tag to (including) the element's end tag.

  Ex : ***<mytag>hello</mytag>***
- Element can contain text or other xml elements
- XML Elements are extensible and they have relationships.
- XML Elements have simple naming rules.
  - Names can contain letters, numbers, and other characters
  - Names cannot start with a number or punctuation character
  - Names cannot contain spaces
  - Underscore, hyphens and digits can be used for a tag name.
  - A colon can be used in case of referring xml namespace entities.
- XML Elements can also have attributes.
- XML documents can be extended to carry more information.

Attributes in XML

- By the use of attributes we can add the information about the element.
- XML attributes enhance the properties of the elements.

  Ex : ***<book publisher="Tata McGraw Hill"></book>***

  – **Generally Metadata should be stored as attribute and data should be stored as element.**

  There are some limitations in using attributes, over child elements.

  - Attributes cannot contain multiple values but child elements can have multiple values.
  - Attributes cannot contain tree structure but child element can.
  - Attributes are not easily expandable. If you want to change in attribute's vales in future, it may be complicated.
  - Attributes cannot describe structure but child elements can.
  - Attributes are more difficult to be manipulated by program code.
  - Attributes values are not easy to test against a DTD, which is used to define the legal elements of an XML document.

  – In the context of documents, attributes are part of markup, while sub elements are part of the basic document contents.

Well- formed XML

- Any xml that follows all the below rules is said to be a well formed XML.
  1. Must contain at least one element. Must have a root element.
  2. Every start tag must have a corresponding end tag.
  3. All tags must be properly nested.
  4. Tags in xml are case sensitive.
  5. Attribute values must be quoted.
  6. Element names can begin with a char or underscore _. Subsequent characters in the name may include letters, digits, underscores, hyphens, and periods.

Error Types in XML

Fatal Error : It occurs when a document is not well formed (Tags are not closed properly) , or otherwise cannot be processed.

Error : It occurs when an XML document is not valid i.e it contains tags that are not permitted by the DTD , and order of tags does not conform to DTD specifications.

It can also occur if the prolog specified in xml cannot be handled by the parser.

Ex. <?xml version='1.3' encoding="us-ascii'?>

Warning : It is generated when the DTD contains duplicate definitions

They are not necessarily an error , but the document author might like to know about it.