UNIT 1 : INTRODUCTION TO XML

1.1 Characteristic and Use of XML

1.2 XML syntax (Declaration, Tags, elements)

1.3 root element, case sensitivity

1.4 XML document:

- 1.4.1 Document Prolog Section
- 1.4.2 Document element section

1.5 XML declaration and rules of declaration.

Introduction

- XML was developed by Tim Bray in 1996(1.0).
- XML 1.1 was released in 2006 sept 26
- XML is a software- and hardware-independent tool for storing and transporting data.
  - XML stands for eXtensible Markup Language
  - XML is a markup language much like HTML
  - XML was designed to store and transport data
  - XML was designed to be self-descriptive
  - XML is a W3C Recommendation

```
<note>
  <to>SYBCA</to>
  <from>STAFF</from>
  <heading>Reminder</heading>
  <body>
      EXAMS ARE ON THE WAY
</body>
</note>
```

The Difference Between XML and HTML

- XML and HTML were designed with different goals:
  - XML was designed to carry data - with focus on what data is
  - HTML was designed to display data - with focus on how data looks
  - XML tags are not predefined like HTML tags are

XML Does Not Use Predefined Tags

- The XML language has no predefined tags.
- The tags in the example above (like <to> and <from>) are not defined in any XML standard. These tags are "invented" by the author of the XML document.
- HTML works with predefined tags like <p>, <h1>, <table>, etc.
- With XML, the author must define both the tags and the document structure.

XML is Extensible

- Most XML applications will work as expected even if new data is added (or removed).
- Imagine an application designed to display the original version of note.xml (<to> <from> <heading> <body>).
- Then imagine a newer version of note.xml with added <date> and <hour> elements, and a removed <heading>.
- The way XML is constructed, older version of the application can still work.

XML Simplifies Things

- It simplifies data sharing
- It simplifies data transport
- It simplifies platform changes
- It simplifies data availability

1.1 Characteristic and Use of XML

- **XML is extensible** – XML allows you to create your own self-descriptive tags, or language, that suits your application.
- **XML carries the data, does not present it** – XML allows you to store the data irrespective of how it will be presented.
- **XML is a public standard** – XML was developed by an organization called the World Wide Web Consortium (W3C) and is available as an open standard

1.2 XML syntax (Declaration, Tags, elements)

- The syntax rules of XML are very simple and logical. The rules are easy to learn, and easy to use.

1. XML Documents Must Have a Root Element

- XML documents must contain one **root** element that is the **parent** of all other elements:
- <root>
  <child>
  <subchild1>.....</subchild1>
  <subchild2>.....</subchild2>
  </child>
</root>

2. The XML Prolog

- This line is called the XML **prolog**:
- <?xml version="1.0" encoding="UTF-8"?>
  - The XML prolog is optional. If it exists, it must come first in the document.
  - XML documents can contain international characters, like Norwegian øæå or French êèé.
  - To avoid errors, you should specify the encoding used, or save your XML files as UTF-8.
  - UTF-8 is the default character encoding for XML documents.

3. All XML Elements Must Have a Closing Tag

- It is illegal to omit the closing tag.
- All elements **must** have a closing tag.

4. XML Tags are Case Sensitive

- XML tags are case sensitive. The tag <Letter> is different from the tag <letter>.
- Opening and closing tags must be written with the same case:
- <message>This is correct</message>

5. XML Elements Must be Properly Nested

- Wrong:
  - This text is bold and italic
- Right:
  - This text is bold and italic

1.3 root element, case sensitivity

1.4 XML document:

1.4.1 Document Prolog Section
1.4.2 Document element section

1.4.1 Document Prolog Section

- This line is called the XML **prolog**:
- <?xml version="1.0" encoding="UTF-8"?>
- The XML prolog is optional. If it exists, it must come first in the document.
- XML documents can contain international characters, like Norwegian øæå or French êèé.
- To avoid errors, you should specify the encoding used, or save your XML files as UTF-8.
- UTF-8 is the default character encoding for XML documents.

1.4.2 Document element section

- An XML element is everything from (including) the element's start tag to (including) the element's end tag.
  - <price>29.99</price>
- An element can contain:
  - text
  - attributes
  - other elements
  - or a mix of the above

- **XML elements** can be defined as building blocks of an XML. Elements can behave as containers to hold text, elements, attributes, media objects or all of these.
- Syntax :

  <element-name attribute1 attribute2>

  ….content

  </element-name>

1.5 XML declaration and rules of declaration

- XML declaration contains details that prepare an XML processor to parse the XML document. It is optional, but when used, it must appear in the first line of the XML document.
- Syntax:

```
<?xml
   version = "version_number"
   encoding = "encoding_declaration"
   standalone = "standalone_status"
?>
```

Parameter	***Parameter_value***	***Parameter_description***
Version	1.0	Specifies the version of the XML standard used.
Encoding	UTF-8, UTF-16, ISO-10646-UCS-2, ISO-10646-UCS-4, ISO-8859-1 to ISO-8859-9, ISO-2022-JP, Shift_JIS, EUC-JP	It defines the character encoding used in the document. UTF-8 is the default encoding used.
Standalone	yes or no	It informs the parser whether the document relies on the information from an external source, such as external document type definition (DTD), for its content. The default value is set to *no*. Setting it to *yes* tells the processor there are no external declarations required for parsing the document.

Rules:

- If the XML declaration is present in the XML, it must be placed as the first line in the XML document.
- If the XML declaration is included, it must contain version number attribute.
- The Parameter names and values are case-sensitive.
- The names are always in lower case.
- The order of placing the parameters is important. The correct order is: *version, encoding and standalone.*
- Either single or double quotes may be used.
- The XML declaration has no closing tag