XML

André Restivo

Index

Introduction   XML   Valid XML   Namespaces   Technologies

Applications

Introduction

Markup Languages

- A **markup language** is a set of words and symbols for describing the identity or function of the component parts of a document.

- Programs can use markup with a **stylesheet** to transform the document into output for screen, print, audio, video, Braille, or reprocessable data formats.
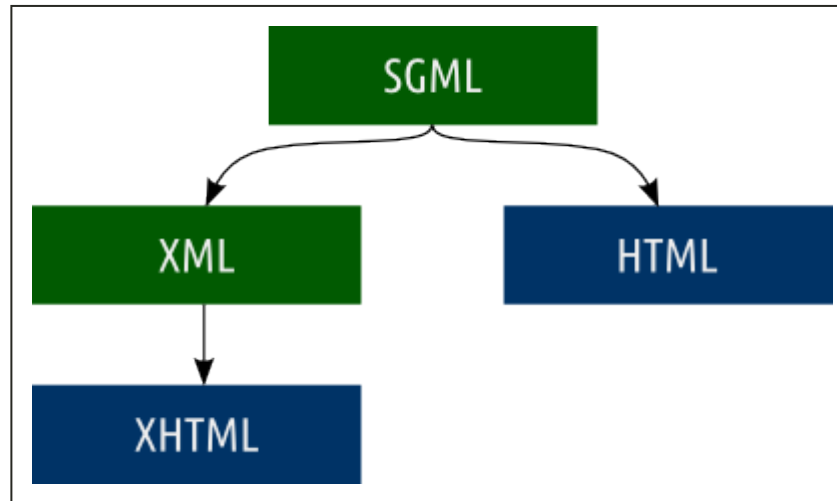
XML

- Extensible Markup Language (XML)

- A **markup language** that defines a set of rules for encoding documents in a format which is both human-readable and machine-readable.

- It is **extensible** because it is not a fixed format like HTML (which is a single, predefined markup language).

- XML is a metalanguage which lets you design your own markup languages for limitless different types of documents.

SGML

- SGML is the Standard Generalized Markup Language, the international standard for defining markup to describe the structure of different types of electronic documents.
- SGML is very **large**, **powerful**, and **complex**.
- XML is a **lightweight** cut-down version of SGML.

SGML, XML and HTML

- SGML and XML are **metalanguages**. They allow users to develop their own languages.
- HTML and XHTML are **concrete** languages with a fixed format.

XML

Well Formed

An XML document is considered well formed if it:

- contains **one** or **more** elements.
- it has exactly **one root**.
- elements **nest properly** with each other.

Processing Instructions

Processing instructions allow documents to contain instructions for applications

Since **XML 1.1**, all XML documents must start with a processing instruction (prolog) indicating the XML version. If not, the document is considered to be **XML 1.0**.

```
<?xml version="1.1" encoding="utf-8"?>
```

The encoding is **utf-8** by default.

Comments

Comments start with a `<!--` and end with `-->`.

```
<!-- This is a comment -->
```

Comments cannot contain double hyphens (`--`).

CDATA

CDATA sections are used to escape blocks of text containing characters which would otherwise be recognized as markup.

They begin with the string `<![CDATA[` and end with the string `]]>`.

```
<![CDATA[
  <warning>These tags are not markup</warning>
]]>
```
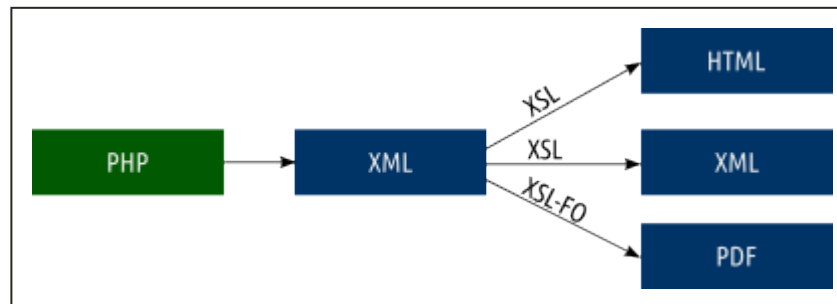

Elements

- Elements are defined by a **start tag** and an **end tag**.
- All elements **must be closed**

Technologies

XSL(t) and XSL-FO

- XSLT (Extensible Stylesheet Language Transformations) is a language for transforming XML documents into other XML documents.

- XSL-FO (XSL Formatting Objects) is a markup language for XML document formatting which is most often used to generate PDFs.

XPath

A query language for **selecting** nodes from an XML document.

Used in several other technologies like **XSL** and **XSD**.

XQuery

A **query** and **functional** programming language that is designed to query and transform collections of structured and unstructured data, usually in the form of **XML**.

Applications

Applications

- XHTML (a XML variant of HTML)
- CML - Chemical Markup Language
- MathML - Mathematical Markup Language
- RDF - Resource Description Framework
- GraphML - File Format for Graphs
- MusicXML - Digital Sheet Music
- SVG - Scalable Vector Graphics
- OSD - Open Software Description
- SOAP - Simple Object Access Protocol
- WSDL - Web Service Description Language
- UDDI - Universal Description Discovery and Integration