XML

XML (eXtensible Markup Language) is a markup language for data. The strength of XML is that you can extend it. XML is the foundation of many Web technologies such as XHTML, AJAX and many other APIs.

XML has a **tag-based syntax** similar to HTML.

What is XML Used For?

XML's purpose is to take **information and add structure to it**. XML was designed to be used over the internet. XML can enable communication between different systems, that may have never been possible before.

Describing Things in XML

The following line must appear at the top of every XML file.

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

Here is an example of what a typical XML file may look like:

```
<Student>
    <name>Regan Koopmans</name>
```

```
    <studentNumber>15043143</studentNumber>
    <degree year="2">Computer Science</degree>
</Student>
```

- Binary data (true | false) should exist as an attribute
- Information that can be subdivided should be an element.

"Valid" XML

An XML document is said to be *well formed* if it conforms to general rules and basic syntactical rules

But we can set custom rules in addition to standard parsing, to set logical rules for how we want to restrict the data. This can be done in two ways:

- DTD - Document type definitions

  > Simple but not powerful. Written in external language.

- XML Schema

  > Written in XML and more flexible and powerful.

DTD

A DTD file can be private/local or public.

To connect a private DTD file we can use the following statement:

```
<!DOCTYPE outermostElement SYSTEM "filename.dtd">
```

Elements

```
<!ELEMENT element>
```

Attributes

Attributes can have one of the following three types:

- #REQUIRED
- #FIXED

- #IMPLIED

Data is in two broad categories

- **#PCDATA** - which is parsable character data.
- **#CDATA** - which is not parsable and therefore forbits any further child elements.

XML Schema

This Schema defines a basic email data structure.

```xml
<?xml version="1.0" enoding="ISO-8859-1"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">


<xs:element name="note">
    <xs:complexType>
        <xs:sequence>
            <xs:element name="to" type="xs:string"/>
            <xs:element name="from" type="xs:string"/>
            <xs:element name="heading" type="xs:string"/>
            <xs:element name="body" type="xs:string"/>
        </xs:sequence>
    </xs:complexType>
</xs:element>
</xs:schema>
```

Advantages of XML

- Content is **separate from presentation.**
- An open format → people can work towards a standard.
- Used on client and server
- Widely supported
- **Disparate systems** can communicate.

Disadvantages of XML

- Not suitable for large amounts of data, **cannot replace databases**.
- Task or technology (such as Javascript) may make JSON more appropriate.
- Images cannot be represented easily.

- Becomes **hard to read** quickly.

Dumping XML into a Database?