The W3C XML Schema

Author: Mihai Rotaru

Date: 1 Dec 2011

XML is designed to be extensible; an XML file can be used to store almost any kind of data. However, XML documents would be of little use without a way of telling applications exactly what kind of data they contain; also, applications would lack a way of specifying with as little ambiguity as possible what kind of data they require.

The first attempt at addressing this issue is represented by the DTD - Document Type Definition. But the scope of XML applications has increased tremendously since its inception. XML was originally used mainly for describing relatively simple documents, such as books, technical reports, articles or web pages. In this context, DTD has proven it's value and is still widely used today; it is the right tool for this kind of job.[1]

However, XML is now being used to store information of a complexity that has increased orders of magnitude since XMLs early days; one example is Scalable Vector Graphics, SVG; DTDs simply do not have the expressive power required to accurately and effectively describe such complex structures. In this context, the need for a more powerful method of describing XML structures becomes obvious.

Therefore, a number of XML schemas were developed, such as RELAX NG, Schematron and W3C's XML Schema. XML Schema, unlike DTD, has an XML syntax itself, and allows for much greater granularity in specifying accepted structures and their types.

One of the most obvious limitations of DTD is that it does not provide a mechanism for specifying the type of an element. For example, I would like to specify the restriction that `year` elements have a value that is exactly four characters long, and each character is a digit. Further more, only numbers in the range of 1900 to 2012 are acceptable. This cannot be accomplished with DTDs; however, XML Schema provides the necessary constructs:

```
<xsd:simpleType name="my_year">
<xsd:restriction base="xsd:int">
<xsd:minLength value="1900"/>
<xsd:maxLength value="2012"/>
</xsd:restriction>
</xsd:simpleType>
```

XML Schema is much more expressive than what this simple example conveys, allowing users to precisely define types of extraordinary complexity. It also has a number of predefined types, among which `xsd:gYear`, which is almost exactly what this example needs, but not quiet.