XML: Powering the Next-Generation Internet

Irsan Widarto

The explosive growth of the Internet has pushed the urge for a universally accepted, Internet-friendly standard for data representation and sharing. XML is this standard. XML, which is a World Wide Web Consortium (W3C) standard, enables organizations to exchange data across applications and platforms, publish product information, export to multiple output channels, syndicate data, automate processing of data etc. etc. And all of this in a programming language-neutral, lightweight, extensible and versatile manner.

Introduction

In this article we will explain what XML is, how you can use it and in what direction it will evolve. At the end we will give a list of resources to get you started developing and deploying your own XML applications.

What is XML?

XML stands for eXtensible Markup Language. Just as with any other language, XML is a set of rules. In this case the rules define how data within a document can be enhanced with tags (the 'Markup') that describe the data. If, for example, our document is a business card:

CrimeFighters, Inc.
Joe Dalton
joe@crimefighters.com

an XML-ized version of this document could look like this:

```
<businesscard>
```

```
<firstname>Joe</firstname>
<lastname>Dalton</lastname>
  <company>CrimeFighters, Inc.</company>
  <e-mail>joe@crimefighters.com</e-mail>
</businesscard>
```

When we talk about "documents" in the context of XML we do not only mean traditional documents like those used for wordprocessing or publishing. Other forms of more or less structured information like music notation, e-mail messages, vector graphics, e-commerce orders, transaction logs etc. are also considered to be "documents."

To define specific semantics and tag sets for certain document types (e.g. business cards) a Document Type Definition (DTD) or XML Schema[2] is created. The XML Schema for our business card example looks like this:

```
<schema
  targetNS="http://void/businesscard.xsd"
        version "1.0"
  xmlns="http://www.w3.org/1999/XMLSchema"
>
```

---

[2]Initially DTDs were the only way to specify semantics and tag set for a document type. Though DTDs and XML Schemas are compatible technologies, it is expected that XML Schema will take over DTD's role. From this point on we will refer to XML Schema when we actually mean DTD or XML Schema.

```
  <element name="businesscard">
    <type>
      <element name="firstname"
               type="string"/>
      <element name="lastname"
               type="string"/>
      <element name="company"
               type="string"/>
      <element name="e-mail"
               type="string"/>
    </type>
  </element>
</schema>
```

XML's extensibility lies in the ability to dynamically extend an XML Schema without disrupting any already created document or application that processes the documents. If, for example, we would decide to extend the business card Schema with a tag for a person's title, we would not have to re-process, or even worse, re-create, all the old documents. Also, applications processing the business card documents do not have to be changed until the new title attribute is actually used within the application.

Typical uses of XML

Now that we have explained what XML is, we will look at two typical uses of XML. First, we will explain how XML can facilitate data exchange across the Internet. Then we will show you why XML is a useful document storage format. It is by no means said that these are the only two uses of XML: XML is what you make it, so you might have completely different reasons to use XML.

XML facilitates automatic processing of documents across the Internet

Web syndication is the process of exposing content and commerce assets on websites in such a manner that other websites can automatically process it. Using XML documents to store and transport the syndicated content is a logical choice due to the fact that XML documents are precisely defined and both platform- and programming language neutral.

An example of web syndication is electronic news-feed delivery. Companies like Moreover and Netscape use XML to enable easy propagation of news headlines and messages. A typical headline in Rich Site Summary RSS (format), Netscape's XML Schema for news syndication looks like this:

```
<item>
<title>
  Out-of-luck soccer team
  loses to fans
</title>
<link>
  http://c.moreover.com/click/here.pl?r69
</link>
<description>
 Press Association
 May 8 2000 3:19AM ET
</description>
</item>
```

Once posted on a publicly accessible (web-)server any other webserver could download the news headlines, create a customized design and publish it.

Though syndication was originally a news industry concept, it is easy to imagine syndication of content on e-Commerce sites. An electronic components manufacturer could for example 'publish' its components catalogue. Customers can then retrieve and import it in their CAD-system or sourcing database.

XML enables single source/multiple output publishing

Tags within an XML document indicate what a piece of data actually is (e.g. a person's e-mail address). The tags do not indicate how that piece of data should be formatted and presented[3] (e.g. underline an e-mail address and make it clickable). How an XML document is formatted and presented is defined in stylesheets. This allows you to publish the contents of a single XML document in various versions (e.g. different languages, different structures, different granularity etc.) and to various devices (e.g. browsers, WAP devices, printers etc.) by creating multiple stylesheets. The stylesheet language (XSL) and the transformation

---

[3]Though it is of course possible to create an XML Schema which does just that: create tags for formatting and presenting purposes. XHTML, the follow-up to HTML, is such an XML Schema.

process (XSLT) have also been standardized by the W3C.

The creation of manuals for customer appliances is a typical example of single source/multiple output publishing: from one XML document an appliance manufacturer could create a user manual, a maintenance manual, a list of parts for service purposes and maybe even an animation showing the basic functions.

XML is also 'ready-made' for multilanguage publishing: it natively supports Unicode, allowing the encoding of texts in the correct local language and character set.

What is next?

Though XML has already gained wide acceptance within the IT-community, a lot of work still has to be done. In this paragraph we will identify some of the developments in the XML universe that are happening or will take place in the near future.

Standardization of XML Schemas across vertical markets and businesses

As explained, there is no such thing as the XML for documents. XML is a meta-language, anyone can create his or her own Schema for any type of document. There is a rapidly growing number of Schemas though, especially defined by and for specific vertical markets or business.

More software supporting XML

To successfully deploy XML, applications and tools that support XML are needed. Besides the growing number of XML-specific software (XML editors, XML formatting engines etc.), more and more software vendors are supporting XML within their products. Microsoft, for example, has announced support of XML for Office2000, enabling the import of and export to XML documents. Especially noteworthy is the growing number of databases offering storage of XML documents. Currently a large portion (maybe even the majority) of XML documents is stored as text files. It is easy to understand that this is not a sustainable situation once XML documents become the core of mission-critical applications (imagine storing your company's accounting records in a text file instead of a database).

Further enhancements to XML standard

The current XML standard defines the basic set of functionality for markup languages. Enhancements to this standard have been proposed for document linking (Xlink and XPointer), input forms (XForms), querying (XML Query) and remote procedure calls (SOAP).

Conclusion

Will XML deliver on its promise? Yes, we think so: XML is just the right thing at the right time and the right place. More and more companies and organizations are using the Web to communicate and exchange data. This was initially done server-to-browser but an ever-increasing number of communication is server-to-server or server-to-some-device-other-than-your-PC. This requires a standard like XML that is flexible, extensible, versatile and probably most important: widely accepted.

XML starting points

World Wide Web Consortium:
    http://www.w3c.org/XML
The XML Industry Portal:
    http://www.xml.org
XML.com:
    http://www.xml.com
Xmlhack:
    developer news from the XML
    community—http://www.xmlhack.com
IBM developerWorks/XML zone:
    http://www.ibm.com/developer/xml

Author's Biography

Irsan Widarto, VP Marketing & Product Strategy, The Connection Factory b.v. (TCF) has extensive experience in developing software and websites as well as defining Internet business strategies. Before founding TCF in 1996, Irsan was employed by Atag Holding and Uniface Technology and worked in Melbourne as an intern for Shell Australia. Irsan holds a Master's degree in Management & Computer Science from Erasmus University Rotterdam. He can be reached via e-mail at irsan@tcf.nl or by phone at +31 10 4046870.

TCF is currently developing Xhive, a database for XML documents. Xhive is the ideal database for mission-critical applications which require the storage of very large volumes of XML documents. Xhive enables easy, fast and object oriented access to the XML documents or document fragments through a Java API. Visit http://www.xhive.com for more info on Xhive.

TCF is based in Rotterdam, The Netherlands and will open its first US office in July 2000. TCF is a Twinning company since 1999 and an active member of W3C.