XML: WYSIWYG to WYSIWYM

A brief look at XML document authoring

Saqib Ali

It all started with cavemen and their cave drawings. All cave drawings were WYSIWYM (What You See is What You Mean). I mean (no pun intended), if you saw a cave drawing, in which a hunter was chasing a mammoth, it meant that a hunter was chasing a mammoth. There were no two ways to interpret the cave drawings. Then came alphabets and words. With words came plain text or documents. Then came XML/SGML for adding information to a document relating to its structure and/or content. An XML document contains both content (words) and an indication of what role the content plays. It's a markup language that adds meaning to the meaning. With the advent of XML, there became a need for XML editors.

Basic Terminology

DTD

Document Type Definition. An XML DTD, written in EBNF, defines the structure and syntax of an XML document. Essentially, it defines the elements, entities, and the content model. For more info, please see this link (`http://www.xml-dev.com/blog/#64`).

Open Standard DTD

Open Standard DTDs are Document Type Definitions that are publicly available for use with various XML aware applications. An example of an Open Standard DTD is DocBook XML DTD. In contrast, WordML developed by Microsoft, is a proprietary XML DTD. WordML needs to be licensed, in order to be utilized by non-Microsoft applications.

XML

XML (eXtensible Markup Language) is a meta markup language that can be used to create other markup languages. XHTML is an example of markup language created using XML.

XML Schema

Similar to DTD, except it's written in XML, and is also capable of defining data types and putting constraints on the content. For more info, please see this link (`http://www.xml-dev.com/blog/#64`).

XSLT

XSLT is a template, which defines what transformations need to be performed on an XML document. XSLT allows XML data to be shared among various XML aware applications, which don't necessarily use the same XML schema. In the context of document authoring and content creation, an XSLT defines what the formatted output will look like.

Valid

An XML document that conforms to rules as defined by an XML DTD or schema is valid.

Well-formed

An XML document that conforms to the syntax rules of XML is well formed.

Proprietary file formats

The popular way of authoring a document is to use WYSIWYG tools (e.g. MS Word or Frame Maker). Documents created using these applications only have formatting information, but lack semantic information (i.e. information related to the structure of the information contained within). Since the formatting information is proprietary to each vendor and the file format is binary, documents created using these applications are locked into the file format of the application's vendor. If you wish to switch word processors (e.g. Microsoft Word to Corel's WordPerfect), you have to go through the painful process of converting all of your documents to the new proprietary format, and often lose valuable data during the conversion process.

Enter XML based document authoring

In contrast, XML allows a document author to create content in a "presentation neutral" format that captures the semantics of the content, rather than the presentational format. XML introduced a way of authoring documents that is completely free from the limitations of old fashioned, non-interchangeable, binary file formats (e.g. DOC, XLS etc). Open Standard XML Document Type Definitions (DTDs) and XML Schemas allow the document authors to create consistent, structured documents regardless of the XML editor they are using. A true XML editor supports any XML DTDs or XML Schema. Thus, the document author is able to move from one XML editor to another, without the fear of the content being locked into one proprietary file format. This is good for the document author, but not good for the software vendor who developed the editor. The only way for vendors to compete for customers is to offer editors that provide the more editing features. If they don't, the customer will just move to a different editor, without losing any content. This has paved the way for the XML editor wars, where each vendor wants to develop an editor that has better features than its competitors' editors have.

Features offered by XML editors

Despite the fact that XML is the Holy Grail for publishers and document authors, nobody wants to use Vi or Notepad to create an XML document. They want tools that'll make XML editing easy and painless. So, the paradigm of WYSIWYG and WYSIWYM was introduced to the XML world.

Essentially the document authors want XML editors that can:

1. Highlight syntax errors
2. Automatically complete XML Tags
3. Indent properly
4. Check for validity against an XML Schema or a DTD
5. Check for XML well-formedness
6. Allow viewing and editing of XML documents in a tree view, and
7. Fix the kitchen sink

Various types of XML editors

The positive outcome of the XML editor wars is that we currently have large number of very good and feature-rich XML editors to choose from. These editors can be divided into three categories: WYSIWYM, WYSIWYG and Text-Based.

WYSIWYM

What You See Is What You Mean is a paradigm that is related to the XML editors, which accurately display the information that is trying to be conveyed, rather than the actual formatting. Since XML doesn't define the actual formatting of the content, these editors are very useful in visually creating and managing the data. A WYSIWYM editor allows a document author to edit an XML document by interacting with a feedback text, generated by the editor. This presents both the knowledge already defined and the options (tags, elements, attributes, etc.) for extending and modifying it. Thus, a WYSIWYM XML editor alleviates the need for the document author to memorize all of the tags, elements and attributes of XML DTD or Schema. WYSIWYM XML editors are also known as *semantical editors*.

Butterfly XML is a powerful WYSIWYM editor, freely available under GNU Public License. In addition to the features mentioned above, it is capable of automatically generating an XML Schema or a DTD based on XML file.

WYSIWYG

What You See Is What You Get is a paradigm that is related to word processing and publishing. It refers to editing

Fig. 1: XML Editors

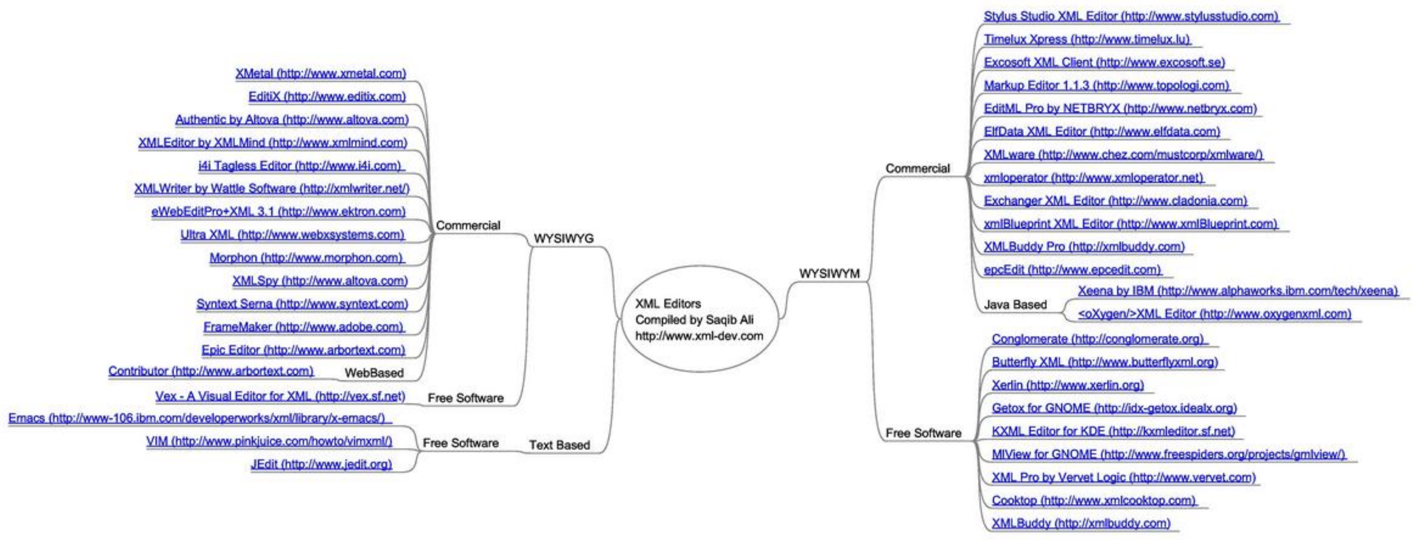

XML Editors
Compiled by Saqib Ali
http://www.xml-dev.com
WYSIWYG
Commercial
XMetal (http://www.xmetal.com)
EditiX (http://www.editix.com)
Authentic by Altova (http://www.altova.com)
XMLEditor by XMLMind (http://www.xmlmind.com)
i4i Tagless Editor (http://www.i4i.com)
XMLWriter by Wattle Software (http://xmlwriter.net/)
eWebEditPro+XML 3.1 (http://www.ektron.com)
Ultra XML (http://www.webxsystems.com)
Morphon (http://www.morphon.com)
XMLSpy (http://www.altova.com)
Syntext Serna (http://www.syntext.com)
FrameMaker (http://www.adobe.com)
Epic Editor (http://www.arbortext.com)
WebBased
Contributor (http://www.arbortext.com)
Free Software
Vex - A Visual Editor for XML (http://vex.sf.net)
Text Based
Free Software
Emacs (http://www-106.ibm.com/developerworks/xml/library/x-emacs/)
VIM (http://www.pinkjuice.com/howto/vimxml/)
JEdit (http://www.jedit.org)
WYSIWYM
Commercial
Stylus Studio XML Editor (http://www.stylusstudio.com)
Timelux Xpress (http://www.timelux.lu)
Excosoft XML Client (http://www.excosoft.se)
Markup Editor 1.1.3 (http://www.topologi.com)
EditML Pro by NETBRYX (http://www.netbryx.com)
ElfData XML Editor (http://www.elfdata.com)
XMLware (http://www.chez.com/mustcorp/xmlware/)
xmloperator (http://www.xmloperator.net)
Exchanger XML Editor (http://www.cladonia.com)
xmlBlueprint XML Editor (http://www.xmlBlueprint.com)
XMLBuddy Pro (http://xmlbuddy.com)
epcEdit (http://www.epcedit.com)
Java Based
Xeena by IBM (http://www.alphaworks.ibm.com/tech/xeena)
<oXygen/>XML Editor (http://www.oxygenxml.com)
Free Software
Conglomerate (http://conglomerate.org)
Butterfly XML (http://www.butterflyxml.org)
Xerlin (http://www.xerlin.org)
Getox for GNOME (http://idx-getox.idealx.org)
KXML Editor for KDE (http://kxmleditor.sf.net)
MlView for GNOME (http://www.freespiders.org/projects/gmlview/)
XML Pro by Vervet Logic (http://www.vervet.com)
Cooktop (http://www.xmlcooktop.com)
XMLBuddy (http://xmlbuddy.com)

Fig. 2: Butterfly XML is capable of presenting the elements that are available at the location of the cursor as defined by the DTD or the Schema

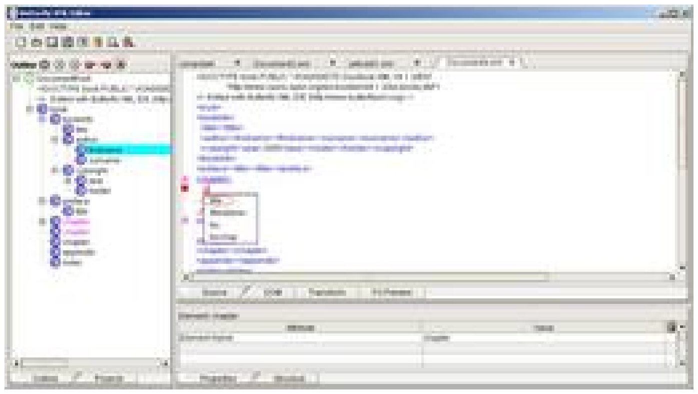
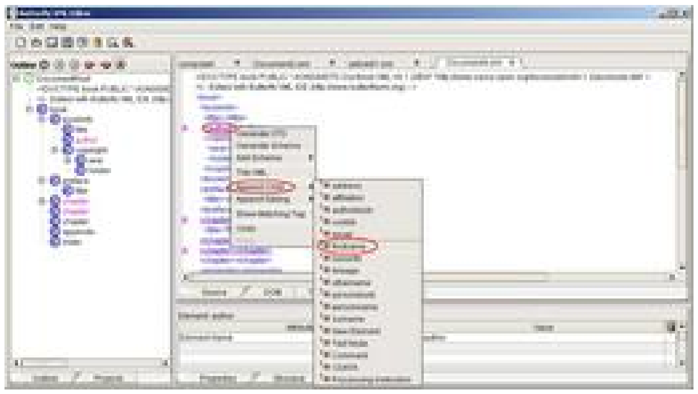


Fig. 3: Right clicking on any tag will display all the child and sibling tags that are available as defined in the XML Schema or DTD.



software that makes sure that the image seen on the screen closely corresponds to the final formatted output. That does *not* mean that WYSIWYG XML editors exclude the semantics of the XML content. All it means is that a WYSIWYG XML editor allows the document author to edit the XML content with the XSLT applied to the content. So the document author doesn't see the XML elements and attributes, instead he/she sees the formatted output as defined by the XSLT.

Vex (Visual Editor XML) is a very capable WYSIWYG editor that uses CSS to provide authors with an MS Word like user interface to create and edit document. It is well suited for ”document-style” documents such as DocBook and XHTML, but is not the best choice for data collection.

Text Based

They are most rudimentary type of XML editors, where the document's author has to manually type in all of the tags and attributes. However, these editors usually check for good structure and sometimes for validity as well. Both Vi and Emacs can be configured to act as a powerful text-based XML editor. The following documentation can be used to configure any of these free editors so they can be used as an XML editor:

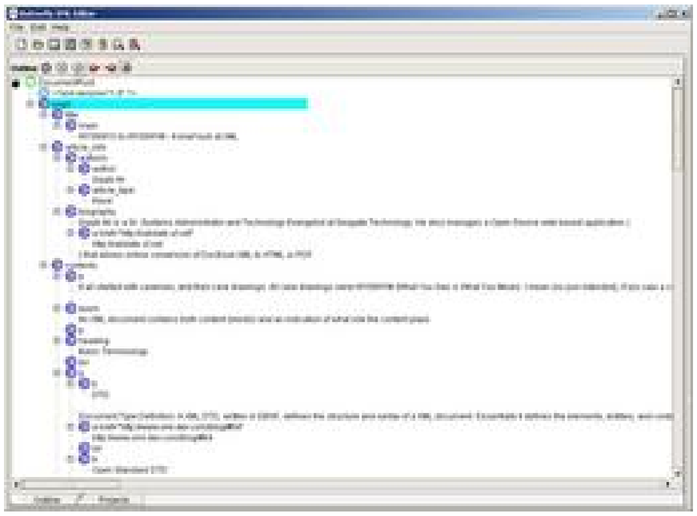


Fig. 4: ButterflyXML offers a Tree View of the XML document. A tree view helps the author to better visualize the semantics of the content.

- Using Emacs for XML documents (http://www-106.ibm.com/developerworks/xml/library/x-emacs/)
- Using Vim as XML Editor (http://www.pinkjuice.com/howto/vimxml/)

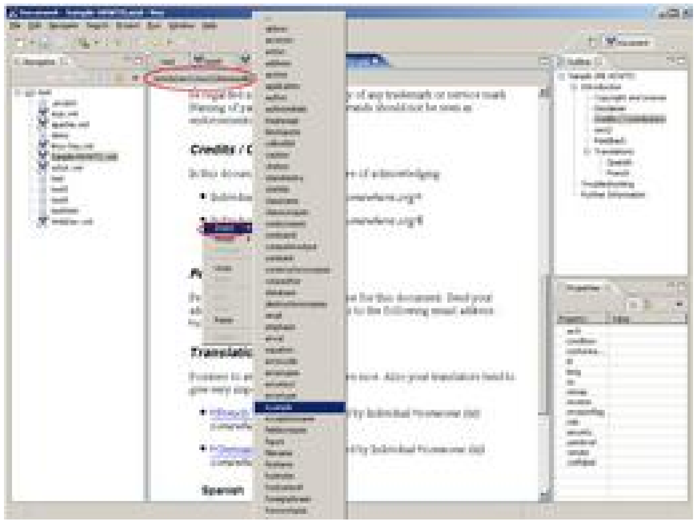


Fig. 5: Like ButterflyXML, Vex is also capable of listing the elements that are valid and inserting them at the current caret location. Available attributes for each element are displayed as well.

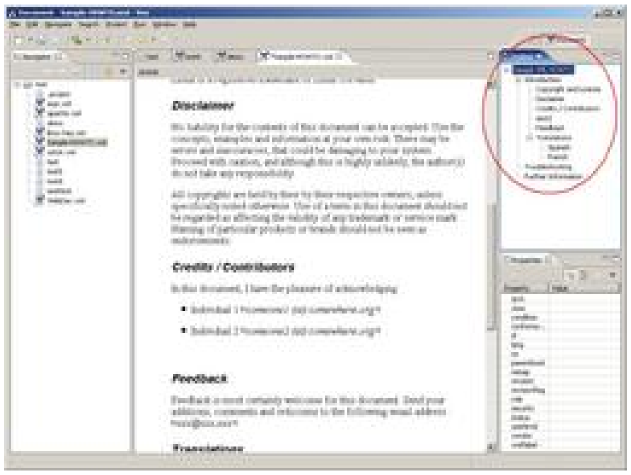


Fig. 6: Unlike ButterflyXML, Vex does NOT provide a tree view of the whole XML document, however it does provide an outline of the document. Outlines provide a good, high-level view of the semantics of the document.

Copyright information

XML allows a document author to create content in a presentation-neutral form that captures the semantics of the content, rather than the presentational format

About the author

Saqib Ali is a Snr. Systems Administrator and Technology Evangelist at Seagate Technology. He also manages a free software web based application (http://validate.sf.net) that allows online conversion of DocBook XML to HTML or PDF. Saqib is also a active contributor to The Linux Documentation Project