

Lecture 11: Current Applications of XML Technology

1. Lecture Contents

This lecture will start by reviewing the wide range of XML technologies available to the software engineer. It will then go on to look at some examples of the many application areas existing today..

2. XML Technologies

DOM	Document Object Model	Language Neutral API Tree Oriented Sees XML as a set of nested objects
MathML		Allows maths expressions to be embedded in documents.
Namespaces		Allows same name to be used for different meanings: Eg <table> as furniture or as formatting Eg <title> as CD title or XHTML tag
RDDL	Resource Directory Description Language	Provides a package of information about some namespace. Describes schemas, style-sheets, and executable code designed to process mark-up from that namespace.
RDF	Resource Description Framework	For describing metadata in, eg catalogues and library cards.
Schemas (XSD)		Describes permissible contents of document. XSD and DTD allow the validation of XML documents. Unlike DTD, XSD is itself validatable XML.

SAX	Simple API for XML	Event driven
SVG	Scalable Vector Graphics	Encoding of line art. Allows embedding of diagrams. Non-proprietary equivalent for Flash.
XHTML	eXtensible HyperText Markup Language	Can be extended using other XML technologies, like SVG or MathML. Useful as a document output format.
XInclude		A way of using parts of many XML documents to form a composite one.
XLink		XML-based hyperlink system. Bidirectional (unlike HTML)
XPath		Non-XML syntax. Used by XSL and XPointer. Identifies parts of a document.
XPointer		Selects part of an XML document. Often used with XLink. Based on URI.
XQuery		Queries across all kinds of XML data. XQuery has a non-XML query syntax. XQuery is an extension of XPath. Operates on structure (data model) of document, rather than the surface syntax.
XSLT (& XSL)	XML Style-sheet Language (Transformation)	Usually output in XHTML. Allows information to be extracted from a document. Allows alternative forms of presentation.
XSL-FO	XML Stylesheet Formatting Objects	Used by publishing industry to format pages. Usually XML (via XSL) to XSL-FO, then to PDF or TEX for printing.

3. Some XML Application Areas

- MusicXML
- DocBook
- SOAP
- CML – Chemical Mark-up Language
- CMS - Content Management Systems
- UBL – Universal Business Language

3.1 MusicXML

The problem of data file incompatibility exists in computer music, as in all other fields. The mechanism behind the problem is a familiar one: a company would set out to create music editing software, perhaps aimed at the markets of school music, or church music, or even theatrical work. They would see no reason not to make up their own standards as they went along. An advantage of an incompatible, proprietary file standard would be that it would be difficult for customers to switch to the products of competitors. Nowadays, however, the situation is the reverse: file incompatibility is a good reason not to buy the product.

MusicXML is not the only competing music file standard. There is:

- NIFF (Notation Interchange File Format), a binary format
- MIDI (Musical Instrument Digital Interface), the current de-facto standard (binary)
- SMDL - Standard Music Description Language (in SGML, an ancestor of XML)
- WEDELMUSIC is an XML compliant format

...and many, many more!

MusicXML is supported by Recordare, who make scanner music recognition software. The idea is that you can put a piece of music onto a scanner, scan it in, convert it from an image to music data and play it on the computer. A typical task might be to scan in a piece of music, change its key signature and print it out again.

The problem is getting recognised (converted) music from the scanning software, into a music editor. That's where MusicXML comes in.

Those who have seen orchestral music will recall that there are two ways of interpreting it. The individual musicians would prefer to have just their own part to work from. The conductor prefers a score, where all parts may be viewed in parallel, bar by bar. There is an open source XSLT style-sheet to convert part-by-part to bar-by-bar, and another to convert in the opposite direction.

Recordare's web site carries the MusicXML definition and information on packages which can read it and write it. Some of these packages are freeware.

3.2 DocBook

The printing industry has always been an enthusiastic user of XML technologies. DocBook is a very popular set of tags for describing books, articles and other prose documents. It is particularly popular for technical documentation.

The standard was drafted in 1991, under the benign influence of Novell and Sun. It is claimed to be an unambiguous semantic and structural mark-up convention. Thus it carries more searchable data than its html equivalent. Documents become searchable databases

3.3 SOAP - Simple Object Access Protocol

SOAP is intended to solve the problems of distributed applications. It is an XML syntax for exchanging messages. A convention for representing remote procedure calls and responses, it is language and platform independent. Microsoft have supported SOAP to the extent that it is a fundamental part of .NET. Nevertheless, a specification has been submitted to W3C, as it is hoped to turn it into an open-source industry standard.

The SOAP XML protocol consists of three parts: Firstly, an envelope which defines a framework for describing what is in a message and how to process it. Then it is a set of encoding rules for expressing instances of data types. Lastly, it is a convention for representing remote procedure calls and responses. The XML of the message is in three parts too, envelope, header and body

3.4 CML - Chemical Mark-up Language

Chemistry was difficult enough at school. The difference between that and real life is that the number of different chemical compounds in existence is far more than a human brain can cope with. The obvious answer is a record system or on-line encyclopaedia. The next problem is that one compound may be known by

many names. Even describing the compound may be done in a number of ways. CML - Chemical Mark-up Language is a way of describing the structure of a compound in XML, designed to be understood by the average chemist. CML can hold 2-D or 3-D molecular information in a variety of ways.

It concentrates on molecules (discrete entities representatable by a formula) and enables the expression of data (as XML elements) such as:

- angle
- atom
- bond
- crystal
- electron

CML is claimed to be the primary application of XML to molecular science. Indeed, the University of California has adopted CML as the chemical technology for its new terascale information and computing grid portals.

CML is NOT in the public domain. It is protected (and therefore stabilised) by copyright and trademarks. Nevertheless, most CML software is Open Source

Namespaces ensure that all XML technologies can interoperate with one another. CML will cheerfully work with other data types, such as SVG, in the same document.

CML files/documents may be generated by:

- Graphical authoring (WYSIWYG) tools available/)
- Editing existing CML documents
- Merging existing CML documents
- Conversion of existing file types
- Direct output from programs (e.g. calculations or instrumental data)
- Manual authoring as a last resort

3.5 Content Management Systems

The term "Content Management System" can mean a number of different things:

- Transactional content management system (T-CMS) to manage e-commerce transactions.
- Integrated content management system (I-CMS) to manage enterprise documents and content.
- Publications management system (P-CMS) to manage publications (manuals, books, help, guidelines, references)

- Learning management system (L-CMS) to manage the web-based learning content life cycle
- Enterprise Content Management Systems (E-CMS) can support web or paper publications, transactions, or customer relationship management content.

Unless otherwise specified, people usually assume that a “Content Management System” is a Web CMS. This is usually a database to store content, combined with some server scripting for delivery.

All CMS feature the separation of content from delivery. This help to minimise the expensive skills needed to maintain some web delivery systems. Thus a content contributor can be a non-technical user. The content kept, maintained and manipulated in the database can be text, images, movies... Page design is separated from content authoring; A page designer is typically a visual editor with page layout training.

CMS enables Content Targeting, i.e. the tailoring of delivery to a client’s profile (eg, geographical location). Dynamic Delivery is the generation of a web page at the time of request

XML can be involved in CMS at a number of possible depths: It could simply be a bolt-on onto a non-XML system, so that it can also handle XML content. The next depth is the ability to do XML storage and handling. The deepest level is native XML technology throughout the implementation and data structures. Organisations usually evolve down this list, because the handling of data in large quantities can be overwhelming, and XML has so many neat answers to this problem.

Often CMS enables and encourages collaboration, i.e. it helps the integration of work on a single project by several concurrent users at separated workstations. This mode of working was pioneered by Lotus (cf: Lotus Notes). The idea of collaborative working is alien to some corporate cultures. Many American firms encourage competition and the concept of collaboration is not compatible. There is a 4th year module on Computer Supported Collaborative Working.

3.6 UBL - Universal Business Language

This module has touched on the reason for XML - a common data storage convention so that applications can talk to each other. Thus a company could buy in different packages from different manufacturers and they would nevertheless be able to interchange files - unlike the situation today. You can therefore appreciate that commerce would run better if there were a common standard for communicating business information to be exchanged *between* companies. As usual, there were a number of false starts and power struggles. The predecessor of the present front runner was ebXML = Electronic Business XML

Now there is a standard called UBL - Universal Business Language. It was originated and promoted by OASIS - the Organization for the Advancement of Structured Information Standards. In order to encourage the adoption of the standard, OASIS offer a free royalty-free library of standard electronic XML business documents such as:

- Order
- Simple and complex order responses
- Order cancellation
- Dispatch advice (often known as a shipping notice)
- Receipt advice
- Invoice

UBL 1.0 was released as an OASIS Standard in November 2004. It is hoped that it would provide an entry point into electronic commerce for small and medium-sized businesses. This would avoid the re-keying of data in fax- and paper-based supply chains, as it is the common business practice at the moment. It is designed to plug directly into existing business, legal, auditing, and records management practices.

At the moment, there are a number of conflicting standards for XML in business. UBL may (or may not) emerge as the winner. It does not really matter which standard becomes the winner; it is only important that one does, and does so as soon as possible.
