XML – A Solution for Publishing Up-to-date Educational Information on the Internet?

Ruairi O'Donnell, Crawford Revie, Monica Landoni, Colm McCartan
Department of Information Science
Strathclyde University
26 Richmond St, Glasgow, G1 1XH, UK
Email: ruairi@dis.strath.ac.uk

The Web was first envisaged as an Information Management tool that helped researcher organise and find interesting documents. However with the popularity of the Internet and the its image as a Publication tool for the masses, we now have a Information Mis-Management Tool. Increasingly people are using it for searching using "Web Portals". Web Surfers are morphing into Web Searchers. XML may help us resurrecting the information chaos into some manageable form. This paper will briefly describe XML and its potential benefits for Educational Delivery.

1 Introduction

The Web is based on technology, which was initially invented for small information spaces with a small number of users. HTTP, URLs and HTML are excellent solutions to a uniform information management system. With the Internet explosion these standards have been stretched to their limits.

The Web has evolved rapidly, so rapidly it seems more like a revolution rather then an evolution. To keep up with these changes the Internet authorities have had a busy time! Each of the original Internet technologies has been extended in numerous ways. HTML has seen the biggest change, initially designed for static text only hypertext pages, now incorporating many different technologies which want to get on the Internet bandwagon.

HTML has obvious limitations that need to be resolved. These limitations include:

- Unpredictable appearance of HTML documents in Web browsers;
- The content of an HTML document is inseparable from the presentation tags;
- Meta information in HTML is restricted to document wide semantic name/value pairs;
- The majority of search engines index only the textual content of HTML pages.

This paper will introduce the main elements of Extensible Markup Language (XML). The paper will outline a case for using XML within Courseware systems.

2 Case Study

The potential of XML to improve the delivery of course material is now illustrated through a short case study. The course under consideration is delivered by the authors to first year undergraduate students at the Business School of the University of Strathclyde, and provides an introduction to Business Computing.

Many of the generic issues relating to courses supported on the WWW are of relevance to this case study, including facilities for off-campus access, student centred learning, standard browser access, etc. However, there are a number of specific issues related to this course that make it particularly suited to web-based support. Firstly, it is a rather large class, by UK standards, with around 700 students. There is a need to give students, particularly in the first year, some sense of contact with teaching staff but it is clearly not feasible to use traditional small group meetings due to the time and labour intensity that this implies. Secondly, the course is delivered by four staff

members from two different departments within the Faculty (it is felt that as this class is for all Faculty students this is a useful mechanism and indeed proposals are in place to increase the number of staff involved). In this context it is important that the students have a 'focal point' for the provision of course information. Finally, students entering this class come with a range of knowledge and skills relating to the use of computers within business. There is therefore a need to tailor material in such a way as to ensure that novices are not over-awed and yet. At the same time, to challenge those with some prior knowledge to extend and in particular deepen their understanding of the subject.

The use of web-based material aids in addressing a number of these issues for example, the web site provides a 'virtual' focal point to which the students can be pointed as a source of initial reference. In addition to finding a course overview, copies of the course handbook and timetable details, information can also be found for classes as they take place each week: tutorial material; preparatory reading; useful web links and even weekly student feedback can be checked on the site. When the 'virtual' solution cannot provide adequate support the site points the student to the most appropriate staff member to deal with their query and encourages e-mail contact in the first instance. The flexibility that the WWW provides to allow the four staff members to interact with the students and the course material, as well as with each other, has many advantages however it can also lead to problems.

The fact that no one person has overall editorial control can lead to conflicting information or to inconsistent presentation over a period of time. Thus there is a tension between the benefits, in terms of flexibility and immediacy, which the WWW offers and the requirement from the students point of view to be able to rely on the site as the authoritative source of course information. The need for increased control on the consistency and accuracy of information has been partially addressed by the creation of course 'style sheets' but this rather informal mechanism does not provide a complete solution and the potential of XML to assist in this context is explored below.

As yet, little progress has been made in addressing the widely differing levels of knowledge students have on entry to the course. This is not because the WWW does not provide suitable mechanisms through which diverse sets of material can be delivered but rather reflects the effort involved in editing and maintaining 3-4 parallel and internally consistent sets of material for the varying levels of student expertise. Once again it is the authors' view that the structuring and delivery of dynamically changing material (i.e. based on student expertise level) can be handled much more adequately using an XML-based approach.

3 XML

XML has generated a lot of attention in the technical press recently where it has been widely referred to as the replacement to HTML. The two languages are, of course, related and share a common parent in the Standard General Markup Language (SGML, ISO-8879). SGML is a complex standard in the form of a meta-language that allows us to describe how a document is structured by means of a Document Type Definition (DTD). A DTD specifies which tags a document may have, what they are composed of and how they are related to one another in the structure of the document in terms of sequence, nesting etc. HTML, for instance, is an SGML DTD defining documents for presentation by a web browser that is 'hard-wired' to understand the HTML DTD. Any users of the web over the last few years will have observed how HTML has been continually updated and expanded to increase its flexibility of presentation, its ability to integrate live data and its functionality for supporting interaction. However, these extensions, as well as being proprietary and non-standard, have been unable to tackle HTML's inability to support document extension or inheritance and its inclusion of document styling alongside structure.

Awareness of the limitations of HTML and the desire to avoid the proliferation of proprietary HTML extensions led to the ratification of XML by the World Wide Web Consortium (W3C) in February of 1998. In terms of complexity, XML lies somewhere between SGML and HTML and is aimed at providing most of the richness of the SGML command set while remaining easy to learn, implement and use. It has been described as a dialect of SGML and is itself a meta-language. Thus, XML allows us to specify not only a marked up document but also to specify which DTD it conforms to. For example, in the case of web documents, there is always an implicit association with the HTML DTD while with an XML document it is usually necessary to refer to the associated DTD explicitly in the header of the document.

This last point highlights what is probably the most important departure from HTML - that XML specifies the structure of a document via its accompanying DTD. Since the document author can define this DTD, any customised tags may be defined and used within the document. The XML markup language is robust enough to describe a wide range of abstract structures in the form of a document and an accompanying DTD describing that abstract structure. XML can therefore be used to describe data objects, structured records, and many other types of structured data. The industry has been quick to see the potential of this new standard for exchanging and sharing not only documents as is currently done on the Internet and various intranets, but also to share a huge variety of structured data using established Internet standards such as HTTP as the mediating protocols.

XML appears to offer a potentially industry-wide, customisable format with the robustness and flexibility to model a huge range of document types. It seems safe to suppose that support for XML in the form of toolsets and application suites will be forthcoming from an industry increasingly accustomed to and demanding of interoperability between vendors' products. A long list of "Who's Who" of software companies--including Microsoft, Netscape, Lotus, IBM, and Adobe Systems--have publicly committed to supporting the XML language.[1]

The three most important parts of XML are the Document Type Definition (DTD); the Extensible Style Language (XSL); and the Extensible Link Language (XLL) [Bradly 1998].

DTD - The Document Type Definition specifies the logical structure of a document. It defines the grammar of a document and enables an XML parser to validate a page's use of its tags. The DTD defines the permitted use of each tag and it's associated attributes. The DTD consists of three structures:

- A description of the nesting of hierarchies for document structures (i.e. A Chapter contains Headings and sections, sections contain paragraphs and so on.)
- The sequence in which the structures are permitted (i.e. A Title is before an Abstract and so on)
- Finally the DTD describes if structures are optional and also repeatable. (i.e. A Paragraph must have at least one or more sentences).

XSL - Extensible Style Language is the language used to specify style sheets for XML documents. XSL enables Web browsers to change the presentation of a document. This is similar to the Cascading Style Sheets (CSS) used in HTML 4.0.

XLL - Extensible Link Language supports today's Web links and it also extends links. It will include indirect links, which should resolve dead links and other hyper link problems. XLL is currently under development. XLL will support a number of link mechanisms [WWW3 XLL 98].

4 Internet versioning with HTML

For a corporate Internet to succeed the content needs to be up-to-date. Most companies have data and information stored in vendor specific file formats, such as Microsoft Word, Corel Draw, etc. For these file formats to be stored on the corporate Internet requires exporting to HTML and then FTP or "Published" to a Web server.

HTML is a poor presentation language with little or no structural information. Sadly most HTML file exporting tools remove important semantic information, stored in vendor specific file formats. Vendors are producing new software tools that can export to HTML but these are merely add-on features. For an Internet to succeed vendors must create software that use a Web file format as the primary file format. In time software vendors will become Web centric, until then we have a versioning problem.

Figure 1.1 illustrates the current situation when an employee saves a Microsoft Word document on a Corporate Internet. This two phased saving & publishing to the Web means that there are always two versions of the document which can lead to versioning control problems. HTML files are secondary file stores, much like a printed document,

[1] For more information on XML have a look at http://www.xml.com/

if the original document is updated then the HTML document must also be changed. For an Internet to succeed then the primary file store, on the Web, must be stored for company access.

5 XML Benefits

We propose storing vendor specific file format in XML. The software vendor can add all the extra "proprietary" information within the file format using XML. Then this XML document can be downloaded and interpreted by any XML compliant Web viewer. Thus we have a vendor neutral viewing facility, allowing users to have thin-clients with few installed programs.

Figure 1.2 shows that vendors can publish the DTD for a document and allow many users to view the information in its full glory. The DTD specifies the complete semantic of the document. In its purest form, XML will allow software vendors to create complex document viewers.

Another advantage of XML is that the appearance of the document can change depending on the media, allowing the document creator to harness all features of each medium. An author must customise a document for a single media and hope that it looks well in different media. Potentially, a single XML based document can take advantage of all the features of a specific printer and all the benefits of a Web browser. Also XML documents can be customised to a number of users, a manager may view an executive summary of a project report while the project team views the complete report.

Research by [Rath 98] has highlighted that XML can assist in the Management of Content, where the document content can be maintained by editorial support systems. Also Metadata can be incorporated into the XML documents so that the documents can be easily indexed and searched. Information stored in a database could also be used by XML documents by enhancing the tag definitions for Database elements such as table names and fields.

Data delivered in XML to a browser can "give Java something to do". This would mean XML could convey data and data types to a browser in a standardised form. In the past migrating data between systems has either been impossible, because proprietary software actually prevented batch upload and download, or it has been a tedious process of reducing complex data formats to flat structures like comma-separated variables. In this process there are ample opportunities for sophistication in the data structure to be lost. XML promises to be the lingua franca for data migration

5.1 Standardised DTDs for XML

For XML to succeed the domain specific DTDs must be globally accepted within a particular domain. There is little point developing numerous DTDs that describe Educational Course Material as we could then have incompatibility issues between the different Educational DTDs. For XML to succeed then there must be some central standards organisation that controls the evolution of one single DTD for a particular Domain such as Educational DTDs. In other Domains XML DTDs are under development, these include:

- Mathematics Markup Language (*MML* used for displaying Math symbols and equations etc)
- Chemistry Markup Language (*CML* to describe molecules and their structure etc)
- Electronic Document Interchange (*XML-EDI*)
- Channel Definition Format (*CDF* used for some 'Push' technologies)
- Synchronised Multimedia Integration Language (*SMIL*)
- XML Metadata Interchange (*XMI* used for exchanging software models)

We propose developing a standard DTD for courseware systems.[2]

5.2 Developing a Courseware DTD using Visual Rhetoric and existing Learning Models

We are currently developing a generic Courseware XML DTD that will allow us to create re-usable Courseware systems. To do this we are using a Graphical User Interface model called Visual Rhetoric to define how course

[2] We are interested in developing a Standard XML DTD that can be widely used. If your Organisations is interested in helping develop this Courseware DTD please contact us.

content should be presented. We are also hoping to improve the quality of the information delivery by using existing learning methods to develop XML DTD tags.

A document can be interpreted as a visible representation of a text according to its semantic contents [Southall 89]. The use of different fonts and typographical styles, as well as the introduction of spacing and pagination rules helps to draw attention to selected parts of the text [Southall 92]. Thus the translation into graphical terms of the text rhetoric which results from both the logical structure of the text and its pragmatic component can be called visual rhetoric [Landoni 97]. The idea is to define those parts that are more important for the comprehension of the meaning of the text.

Visual rhetoric provides the reader with a graphical mark up language that is immediately recognisable on the basis of previous reading activity. Different graphical presentations suggest different readings and affect deeply the interpretation of the contents of the same text. These observations lead one to conclude that visual rhetoric is a crucial aspect for both reading and browsing a document. We will be adopting this approach visual rhetoric when developing the XSL and XLL that will define the presentational aspects of the DTD. It is hoped that visual rhetoric will give us the possibility to control/predict the way student interprets and interacts with the Courseware documents.

One of our main goals is to improve we improve the effectiveness of a Courseware system. A DTD that uses 'semantic elements to exploit knowledge inside the information' can help increase the quality of the data and the service of online information [Rath 98]. At present we are identifying teaching models that encapsulate most learning methods and teaching techniques. We will then create XML tag definitions that correspond to each element of these teaching models. It is a relatively straightforward task to translate each of the teaching elements into DTD tag definitions. However developing a DTD from scratch is time consuming, and requires a great deal of testing [Maler 96].

6 Commercial acceptance of XML

Many software vendors are investigating XML as the future file format for universal viewing, editing and saving. "XML can potentially revolutionize the Web," said Chris Lilley, from the World Wide Web Consortium (W3C). "It's not a magic bullet ... but as a cornerstone of an open network computing environment it's very important." Microsoft's plans a new HTML-XML hybrid file format and both Lotus and Corel plan to support XML in their next releases. It could eventually supply the basis for universal file formats usable by all suites [PC World 98].

Berners-Lee [Montgomery 98] is optimistic about the use of XML markup tags outside the Web. This will overcome the problem of backward incompatibility between document formats. The W3C is talking to Microsoft Corporation about using XML as a cross-platform file format for Word. Berners-Lee suggests that HTML will remain reasonably static at Version 4.0, as XML allow vendor to extend and customise tags for specific purposes.

7 Future Work

At present there are very few software tools that support XML. To us XML properly within a educational environment we need software that can support XML: author a course; publish course material; capture and manage ongoing student/instructor interaction; store course material and allow easy retrieval methods for all types of users; and a user interface/browsers to view interact with the material. To date this XML support is limited however most software vendors have identified XML as a feature which will be added to forthcoming products. So to conclude when it comes to XML: "Watch this (Name)Space!".

References

[Landoni, M. 97] *The Visual Book System: A Study of the Use of Visual Rhetoric in the Design of Electronic Books***,** Glasgow: Department of Information Science of the University of Strathclyde (PhD Thesis).

[Southall, R. 89] *Interfaces between the Designer and the Document*, in, André, J., Furuta, R. and Quint, V. (eds.)*Structured documents*. Cambridge: Cambridge University Press. pp. 119-131.

[Southall, R. 92] *Presentation Rules and Rules of Composition in the Formatting of Complex Text*, in, Vanoirbeek, C. and Coray, G. (eds.) Electronic Publishing '92 - International Conference on Electronic Publishing, Document Manipulation, and Typography, Lausanne, Switzerland . Cambridge University Press. pp. 275-290.

[Bradly 1998] *The XML Companion*, N. Bradley, Addison Wesley, ISBN 0-201-342855.

[PC World 98] *File Format Frustrations: New Clouds on the Horizon*, PC World July 1998, http://www.pcworld.com/current_issue/article/0,1212,3929+7+0,00.html

[Maler 96]. Developing SGML DTDs, Maler and Andaloussi, 1996, Prenctine Hall ISBN 0-13-309881-8.

[Montgomery 98] *Berners-Lee Upbeat on XML*, Paul Montgomery, PC Week Australia, 04.15.98, http://www.zdnet.com/pcweek/news/0413/15ewww.html

[XML-spec 98] XML Complete Specification, World Wide Web Consortium, http://www.w3.org/pub/WWW/TR/WD-xml.html

[WWW3 XLL 98] *XML Linking Language (XLink)*, World Wide Web Consortium Working Draft 3-March-1998, http://www.w3.org/TR/1998/WD-xlink-19980303

[Rath 1998] *XML: chance and challenge for online information providers*, Hans Holger Rath, Proceeding of the Online Conference 1998, London.

[Sonnenberg 96] *Design of a Reusable IR Framework*, Gabriele Sonnenberg, Hans-Peter Frei, http://www.ubs.com