

Revolutionizing Content for XML

By Debbie Kenny, Director of Marketing at Information Mapping, Inc.

With promises to revolutionize how information is created, stored, and accessed, all the major Web tools are rushing to support XML, and many organizations are already using XML as the basis for creating sophisticated database-publishing solutions.

If you think of XML as HTML on steroids, you will have an imperfect, possibly even politically incorrect, but useful picture of what all the fuss is about. Organizations that are publishing large quantities of content to their electronic information systems such as Web sites, Intranets, and Extranets, are quickly realizing that HTML has several drawbacks that limit its usefulness for “industrial strength” applications. Perhaps most significantly, unlike HTML, XML allows a publisher to store “meta information” about a piece of content with the content, and use that information to access and use the same piece of content, from a database, in different ways. For example, a description of a product might be tagged in XML, stored in a database, and accessed in different ways for use in a product specification, a Web site, a training manual, *and* a troubleshooting guide. When the original information is updated or revised, it is instantly updated in all of its various forms. Furthermore, unlike HTML using XML allows you to separate the information content from the presentation/layout data. This gives you even more flexibility - and allows you to redesign entire Websites without having to rewrite any information.

In order for XML to be effective for reusing content, source information must be structured in a consistent, standardized fashion, and rules must be defined that describe how the information is structured. Otherwise, there is no way to break the information up in to useful, logical units that can be used and reused in multiple ways. Often, when organizations begin to explore an XML based solution to their publishing needs, they are able to find and implement supporting technologies, but they run in to difficulties when they convert their source information. Many organizations find that the existing information they have has been created in very inconsistent ways. There may be no consistent use of styles, no clear units of information, and no discernable structure to their documents. Software vendors tell them “go back and structure your content, and when you’re ready we will convert it into our system.”

The question from many organizations is “how do we go about structuring our content so it will work with XML?” The answer is that organizations need to adopt an information standard and a method for categorizing, organizing, and presenting all of their content. They must train their content authors to use this method and standard consistently, and the content needs to be reviewed and reworked in order to ensure that it is structured consistently according to the standard.

According to Information Mapping, a consulting firm that specializes in helping clients’ structure content, there are four elements that are critical to structuring content effectively for reuse with XML.

Identifying Information Types

Content should be separated into distinct type of information based on the purpose. Procedures, processes, principles, facts, structures, and concepts each have a different purpose for the audience and are used in different ways. When content is identified and labeled according to its information type, it becomes easier to categorize and manipulate in meaningful ways. Separating source information by information type and including a description of the information type as Meta information, connected with the content is an extremely powerful basis for structuring information.

Creating Modular Content Units

Unlike traditional prose, structured content must be broken into small manageable units that are carefully defined. Well-defined, modular units of content provide an ideal structure for “tagging” information in a meaningful way. Content providers must be trained to create consistent units of information based on well-defined guidelines.

Applying Research-Based Organization Principles

Research has shown that information is most effective when it is chunked into small, manageable, units of relevant, consistent information, and made accessible to users when, and if, they need it. Consistent application of the principles of chunking, labeling, relevance, consistency, accessible detail, integrated graphics, and the hierarchy of chunking and labeling, ensure that content is well structured and consistent enough to be systematically tagged for XML.

Presenting Content Consistently

Information that is tagged in pure XML does not contain inherent formatting. A style sheet, applied at the point of publication, specifies the formatting. Nevertheless, consistent formatting can be critical to the XML conversion process, since the conceptual structure of a document is usually represented by its physical appearance. For example, most XML converters are able to identify and tag all labels of a certain type, size, and weight as the heading for a new part of the document. Using pre-defined style sheets and formatting standards creates a document with an inherent structure that is visually obvious and can often be cleanly converted to XML “as is.”

Organizations that have their content structured prior to converting it into XML have an advantage because their information is already suitable to support the promises of database publishing with XML. Organizations that are willing to invest in their information will benefit from new, more flexible ways to create, manage, maintain, use, and reuse information. In the information age, how well organizations perform these tasks is a key indicator of who will ultimately be successful.

About the Author

Debbie has been with Information Mapping, a leader in knowledge management and Web solutions, since 1987. Currently the Director of Marketing, Debbie has held a variety of roles in the organization including, Product Manager, Business Development Manager, Software Division Manager, MIS Manager, and Production Manager. In these positions, Debbie has played an essential role in integrating technology into Information Mapping's products and services.