

# **XML (Extensible Markup Language)**

## **Introduction**

This paper will discuss XML (Extensible Markup Language). This discussion will include an introduction to XML, the utility of XML and how XML is different from other languages currently in use. The purpose of the paper is to provide the reader with an understanding of this powerful new tool which is likely to supplant the current languages such as HTML (Hypertext Markup Language) and DHTML (Dynamic HTML) within two years.

## **Significance of the topic**

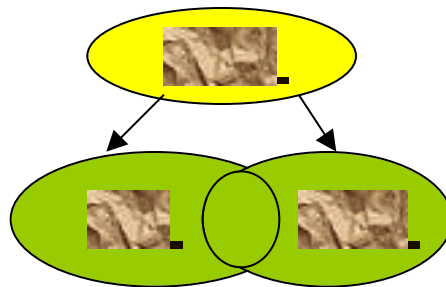
One version of XML has been standardized according to the W3C (World Wide Web Consortium), the organization which oversees such things. According to people in this area, XML is likely to supplant the current languages in use such as HTML. In fact, the new Windows 2000 program is entirely XML based. Since XML can be used across platforms and read by a variety of machines (unlike HTML which can only be read by a browser), it is likely that XML will supplant HTML and DHTML in web design within 5 years. It will also be used to improve the effectiveness of use of large EDI (Electronic Data Information) systems.

## What is XML?

XML is a language for documents identifying structured data in a quite simple way. Structured data includes both content (e.g., words, pictures) and some action indication (markup; tags). For instance, content in a section heading has a different meaning from content in a footnote, which means something different than content in a figure caption or content in a database table. XML documents are text based. Therefore, after creating your document, you can share it with everybody regardless of the computer or operating system s/he uses.

## Where does XML come from?

XML is derived from SGML (Standard Generalized Markup Language), the same parent as HTML. Roughly speaking, XML is a restricted form of SGML.



## How does XML differ from SGML?

SGML has been the standard, vendor-independent way to maintain repositories of structured documentation for more than a decade. It is a complex metalanguage (a language designed for talking about other languages) used to exchange documents. However, it is not well suited to serving documents over the web (for a number of

technical reasons). Because XML comes from SGML, any fully conformant SGML system reads XML documents. However, using and understanding XML documents does not require a system that is capable of understanding the full generality of SGML. XML has 10% of the complexity and 90% of the power of SGML (Tittel and Boumphrey, 2000).

## **How does XML differ from HTML?**

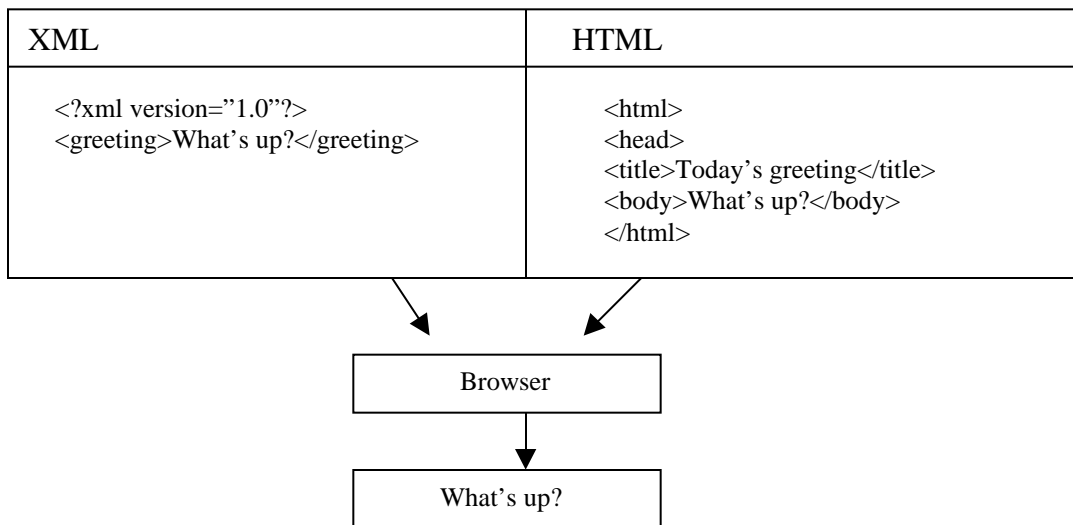
XML documents use the same syntax as HTML pages (e.g., tags, attributes). Although XML and HTML are similar in lineage and construction, they are two very different markup languages. Importantly, XML can solve problems that HTML has.

### HTML Limitations:

- HTML doesn't include the mechanisms for maintaining fine control. A web designer can't specify the display size of a document or control the size of a browser window. Although HTML 4.0 includes <font> tags to help a web designer manipulate font style, size, and color, users can override these settings with their own.
- HTML consists of a closed and predefined tag set. That is, both the tag semantics and the tag set are fixed. For example, <h1> is always a first level heading and the tag <author> is meaningless.
- HTML cannot display data in multiple formats. Therefore, a web designer should try every available browser (e.g., Netscape, Internet Explorer, Mosaic, Hot Java, Mozilla and Opera) and on every platform (e.g. web TV, PDA

(personal data assistant like Palm Pilot), PC, Mac, etc). It is impossible to know exactly which browser and platform are being used to view web pages.

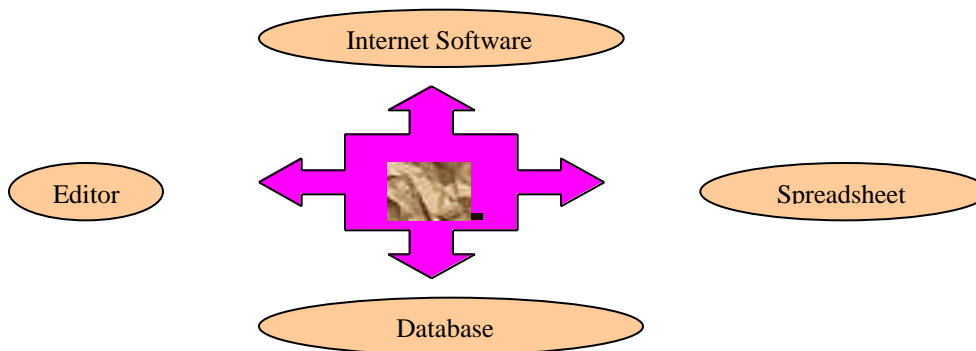
Example comparison of XML and HTML:



**Why XML?**

- Data exchange and database connectivity

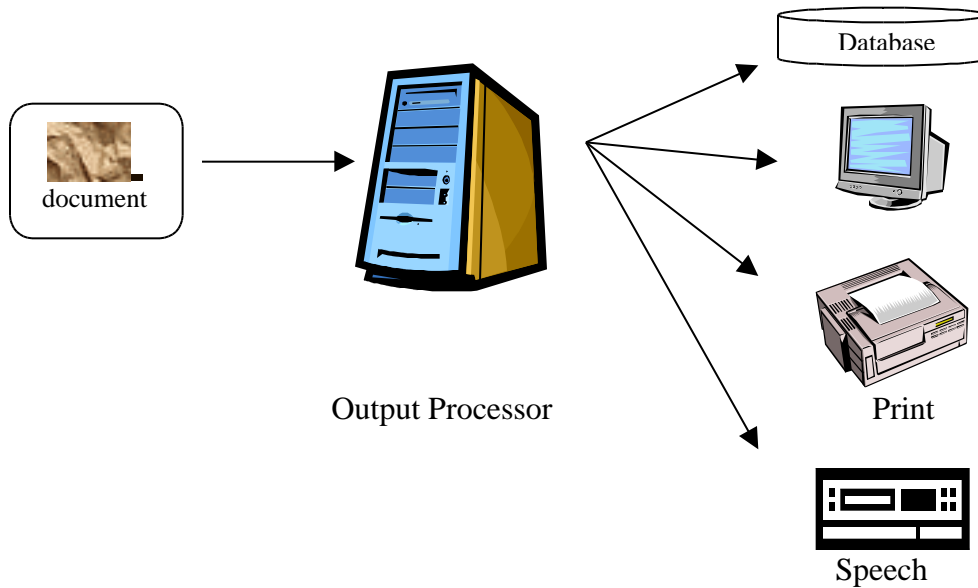
XML will become more important for data exchange over the Internet in the future. It's always difficult to find an interchange format suitable for data transfer between databases from different vendors on different operating systems.



- Multiple Outputs

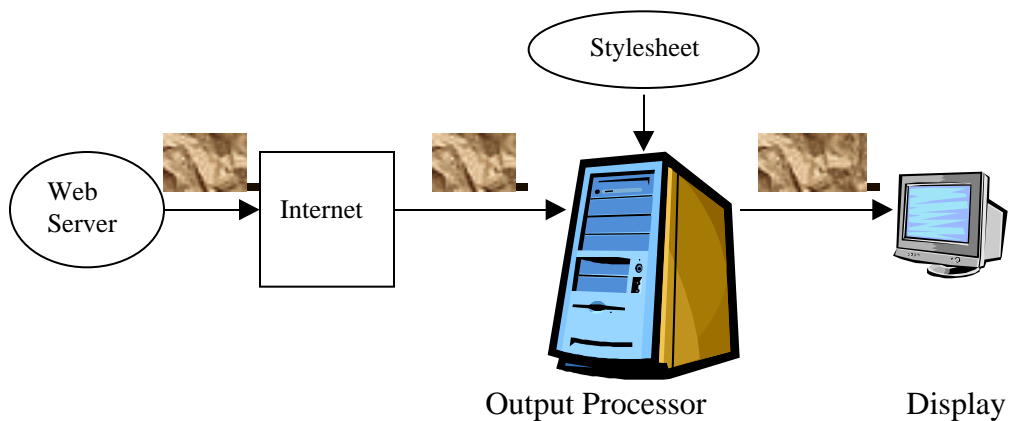
HTML and DHTML can be read only by browsers. This is not true of XML.

Therefore, when you enter an XML document, it can be read and parsed (interpreted) by databases, web displays, printers and sound machines.



- Works well with HTML

At the moment XML can be used with HTML. The stylesheets can be incorporated into the design to give the designer more control over the look of a page.



## **Summary**

XML is a language that separates the data from the presentation of the data. This means that the form and the content of the data are separate, unlike in the current languages. This makes the data more easily understood by a person reading the data since the tags are meaningful. It allows data to be shared across platforms and across agencies and institutions. For example, some industries are developing standards of their own so that any store in that industry will all use the same tags. For instance, say the jewelry industry wanted a huge database to know what every store had in stock. With standard XML this would be possible and all of the stores data could be read in the same way by all of the other stores. There would be no problem sharing information. In this way, XML is a kind of universal translator that immediately lets people of different languages communicate seamlessly with each other.

XML was created so that richly structured documents could be used over the web. The other alternatives such as HTML and SGML are not sufficient for this purpose. XML is a tool that gives designers more control over the look of their pages while rendering data more comprehensible and useful. However, the main problem now is that there is only one browser that is fully compliant—Mozilla. So, it is hard to construct documents in XML for the web. However, that is expected to change shortly.

## **References**

Tittel, E., and Boumphrey, F. 2000. XML for Dummies. IDG Books Worldwide, Inc.

CA.

Walsh, N. 1998 “What is XML?” XML .com web article.

## **Related Links on the Web for the Topic**

<http://www.xml.com/pub/98/10/guide>

<http://wdvl.com>

<http://xmlsoftware.com>

<http://xml.com>

<http://webmonkey.com/authoring/xml>

<http://w3.org/xml>

[http://www.oasis-open.org/cover/thompsonSchemaSlides19991220\\_files/frame.htm](http://www.oasis-open.org/cover/thompsonSchemaSlides19991220_files/frame.htm)

<http://webdeveloper.com>

<http://jayclark.com>

<http://www.stars.com/authoring/languages/xml>

This paper is written by Hye-Yeon Lim for the course EDC 385G Multimedia Authoring at the University  
of Texas - Austin