Markup Languages

Early word processors used "markups". These were instructions embedded in the documents that specified how parts of the text should be presented and included effects like bold type, italics, etc. While this was very useful, these instructions were of little use in helping determine the content or meaning of any parts of the document. In the late 1960's, the utility of the "markup" idea was extended by some researchers (including Charles Goldfarb). These people realized that marking electronic documents with more general tags that indicated the meaning of parts of the documents could be very valuable. For example, many documents have a title, an author, a date, etc. If all of the similar portions of the documents in a collection were marked up in a standard manner, it would be possible to write programs that could locate documents by a certain author, produce profiles of all of the documents, etc. In addition, using the meaningful markup tags, one could print the documents in different styles based on preferences. For example, you could print a document in one font for one person and in another font for another person by specifying at the time of printing that titles should be printed in a certain size, the authors' names printed in italics, etc.

SGML

This concept proved so useful, that SGML was made into an international standard (ISO 8859) in the mid-1980's. While SGML is utilized by thousands of institutions, companies and organization, SGML is a very large, complicated standard and does not lend itself to casual use.

HTML

HTML was first developed by Tim Berners-Lee as a hypertext language for linking information among researchers. His system used a set of uniform addresses to refer to documents on different computers, a set of rules (a protocol) for transmitting the documents and a simple markup language for encoding the documents. He based his HTML on SGML, but used only a small subset of markers or tags. Documents using these tags could be interpreted by programs written for a number of different computer architectures. Because the programs and the computer platforms varied, the documents would appear differently in a program written for a very high-powered computer that could display different fonts, sizes and colors, than they would in a program written for a low-end text-based system. But the important notion was that these documents could be requested from a server, read by almost anyone with a computer and could provide links to other related information.

The World Wide Web spread quicker than anyone expected. People wanted to include more than just text in these documents. They wanted to include other media such as images, icons, text styles, and other media included. The rapid spread of the WWW left standards organizations in the dust. Two major browsers emerged, and gradually each responded to the demand for more complex media and more style in different ways with different tags. This diversion meant that pages

developed for one browser may not be able to be interpreted by another browser. This was in direct opposition to the initial concept of the WWW, which was to provide universal access to information. Today, there are five major browsers, Edge, Firefox, Safari, Opera and Chrome. All five adhear almost identically to each other when it comes to HTML, though legacy IE browsers (IE5, IE6, IE7, IE8) tend to cause the most issues due to their continued widespread use throughout the world.

HTML is still a means of formatting text for display in a browser window. It is not a procedural programming language like C, C++, Java, Pascal, or Fortran. HTML5 is the latest standard version of HTML and is the new standard. HTML5 is a major release and is helping make HTML a powerful tool once again. It is bringing the first updates to HTML forms since the release of HTML2.

XML

XML is another derivative of SGML developed by ISO in the mid1990's. It is called "extensible" because it does not consist of a set of tags like HTML. It is really a "meta" language – a language from which to create other languages. It provides a consistent set of rules for creating these other languages and can be thought of as a lighter version of SGML.

XML was developed by ISO in response to a need for consistency among browser markup languages in response to the desire for a means of using tags that were meaningful in terms of content, not just page appearance. The latter is close to the initial goals of SGML, and, in fact, XML makes it easier to share SGML-style documents over the WWW. In a way, XML can be viewed as an intermediary between HTML and SGML.

XML furnishes a common syntax for the creation of specialized markup languages for any domain or discipline. It is not a procedural programming language like C++, Java or Fortran. It is a means of describing information that will be stored, transmitted to others and processed by a program written to interpret it. There are already specialized markup languages for Mathematics (MathML), Chemistry (CML) and business data (XBRL). With XML, a group can create a markup language for books, games, sports, teams, people, animals, finance, products, services, etc. For example, a book may be described in XML like this:

```
<book>
        <title>Gone with the Wind</title>
        <author>
                <firstname>Margaret</firstname>
                <lastname>Mitchell</lastname>
                <flag gender="F" />
        </author>
        <publisher>Warner Books</publisher>
        <isbn>0446365386</isbn>
        <review>
                Sometimes only remembered for the epic motion picture and
                "Frankly … I don't give a damn," Gone with the
                Wind was initially a compelling and entertaining novel.
                It was the sweeping story of tangled passions and the rare
                courage of a group of people in Atlanta during the time of
                Civil War that brought those cinematic scenes to life. The
                reason the movie became so popular was the strength of its
                characters--Scarlett O'Hara, Rhett Butler, and Ashley Wilkes
                --all created here by the deft hand of Margaret Mitchell, in
                this, her first novel.
        </review>
</book>
```

The XML code shows an element book which has subelements title, author, publisher, isbn and review. The subelement, author, has subelements for the first and last names of the author. It also contains an empty element, flag. The flag element has an attribute, gender which indicates the gender of the author as "M" or "F". Empty elements can either be enclosed by placing the slash at the end of the beginning tag or by explicitly using a closing tag. In other words the following two set of tags are equivalent.

```
<flag gender="F" />

<flag gender="F"></flag>
```

In order to process an XML document, an XML parser is employed. The XML parser locates the tags and comments in an XML document. Programs written in Java, C++ or other languages can then respond to the elements found by the parser. For example, one might write a program that displays a book's title in bold and the author in italics. More complicated programs might also search for other books by this author and provide a link to those records.