

The History of Markup Languages

By Donald J. Zijm

Luminous Solutions LLC

<http://www.luminoussolutions.com/>

Introduction

The roots of Markup languages go far and wide. There have been many people that contributed to the development of past and present Markup languages. In an effort to keep this paper brief, I will only mention some of the major events and people involved.

This paper will first explain what a Markup language is and the value of using a Markup language. In an effort to keep the focus on the history, I have purposely left out examples of the various Markup languages mentioned. For more information on Markup languages please visit the Web Site links included at the end of this paper.

Markup Language Explained

With the invention of the telegraph, we saw the beginning of what is known as the Information Age. This machine was a tool that could transmit information over long distances. Dots and dashes, known as Morse code, is the “language” that is used to convey the message. The telegraph machine doesn’t “know” anything about the information being transmitted. Today, the computer is at the heart of the Information Age. As our demand for information has increased, so has our demand to have the machine “understand” the information and make use of it.

The term Markup has its origin in publishing. The stylistic instructions that were written on a manuscript prior to being typeset are known as Markup. In the computer industry, a Markup language is code that is used to explain both the structure and information in our documents. A very popular Markup language and a prime mover for the World Wide Web is HTML. One that is not as popular but is paving the way for the future is XML. How did these languages evolve into the languages that we know today? Let’s take a look at the brief history of Markup languages and some of the people involved their creation.

Markup Language History

In September of 1967, during a meeting at the Canadian Government Printing Office, William Tunnicliffe gave a presentation on the separation of information content of documents from their format. During this period of his life he was chairman of the Graphic Communications Association (GCA) Composition Committee. At the same time, Stanley Rice, an editor at a Major Publishing house was writing about “Standardized Editorial Structures”. This was the beginning of a movement to separate the formatting of a document from its content.

Building on the works of Tunnicliffe and Rice, the GCA created a committee called the GenCode Committee. Some important concepts that exist today are the product of this committee. In 1969, Charles F. Goldfarb, a graduate from Harvard Law School, built on the work from the GenCode committee and with help from Ed Mosher and Ray Lorie at IBM went on to develop Generalized Markup Language (GML).

Goldfarb felt that GML should both describe the documents structure and be structured in a way such that it could be both human readable and machine-readable. Goldfarb continued his work at IBM as GML began to grow in popularity. Several years later and with the influence of hundreds of people, Standard Generalized Markup Language (SGML) was born. SGML added additional concepts that were not part of the GML project such as link processing, concurrent document types and most importantly the concept of a validating parser that could read and check the accuracy of the Markup.

In the 1970’s the popularity of GML and SGML grew as the benefits became obvious. By 1980, 90 % of IBM’s documents were produced using GML. The US Department of defense began using GML in the 1980’s and eventually the Army required that contractors submit documentation using SGML. SGML was becoming more standardized and, at the same time, was growing as a standard both in the US and internationally.

As Markup languages go, SGML is powerful, flexible and complex. HTML, a subset of SGML is on the other hand easy to learn but not nearly as powerful. The credit for the development of HTML and the World Wide Web is given to Tim Berners-Lee while working at CERN (European Laboratory for Particle Physics) in Geneva. In 1989, Tim Berners-Lee and a colleague, Robert Caillau, were working on a linked information system. Their system used a NeXT computer and incorporated the concept of hyperlinks. Tim realized the need for a Markup language that was easy to use and implement into their system. In 1991 the Web debuted on the internet and it was the simplicity of HTML that made the Web grow at a feverish pace.

In October 1994, Tim Berners-Lee founded the World Wide Web Consortium (W3C). The mission of W3C is to create an open forum in which to lead the technical evolution of the Web. HTML went through several versions as new features were added along the way. The W3C realized that a major drawback of HTML was that it focused on presentation and it combined instructions for presentation in the same document that

contained the content or information. Although HTML was easy to learn, the combined content and presentation instructions made building large sites complicated.

As the Web grew in complexity, a new Markup language was needed to support the evolving demands. A Markup language was needed that was easier to use than SGML but still would be capable of describing the content. In November of 1996, the initial working draft for Extensible Markup Language (XML) was presented.

XML was not meant to replace HTML. The goals of each are different in that XML focuses on the structure of data and what the data is while HTML focuses on how it looks. As the Web continues to mature, many believe that we will see both Markup languages coexist. In addition, XML is quickly becoming a standard way for applications to exchange information. With XML we have a structured, human-readable and machine-readable way to exchange information.

Conclusion

In conclusion, our need for exchanging information in a structured and definable way has led to the creation of Markup languages. The journey from GML to XML has seen the influence of many people. As the information age matures we are likely to see Markup languages playing a central role in the exchange of information.

Links:

<http://w3.org>

<http://www.xmlfiles.com/>

<http://www.w3schools.com/>

<http://xml.coverpages.org/>

<http://www.w3.org/MarkUp/SGML/>

About the Author:

Donald J. Zijm is the founder of Luminous Solutions LLC. Luminous Solutions LLC develops Custom Information Management Software Solutions.

Contact Information:

Luminous Solutions LLC

Email: don@luminoussolutions.com

Web: <http://www.luminoussolutions.com>