ASML: Automatic Site Markup Language

CHARLES B. OWEN  cowen@cs.dartmouth.edu
FILLIA MAKEDON makedon@cs.dartmouth.edu
Dartmouth Experimental Visualization Laboratory, Department of Computer Science, Dartmouth College, Hanover, NH, USA

Abstract. Creation of large and complex World Wide Web sites is hampered by the “page at a time” approach of many tools and the programming knowledge and custom software development required for automated solutions. This paper describes the development of Automatic Site Markup Language (ASML). ASML is a new markup language designed to automate and facilitate the production of large, complicated web sites which can include dynamic content or content that changes with time. ASML extends HTML with new, high-level features while still preserving complete compatibility with common browser and server technologies. It has powerful indexing and searching facilities, and enables the automatic translation of document formats. Most importantly, ASML provides HTML-like features at the site level rather than just the page level.

Keywords: World Wide Web, HTML, SGML, markup languages, multimedia authoring, courseware

1. Introduction

The World Wide Web (WWW or the web) is a component of the Internet which allows for simple and efficient electronic publication of documents. Automatic Site Markup Language (ASML) is a new markup language designed to automate the construction of World Wide Web publications. It centralizes functionality, decreases duplication of effort, and supplants most uses of scripting languages and custom programming in site development. It has been utilized and tested in the creation of several historic sites at the Dartmouth Experimental Visualization Laboratory (DEVLAB) aimed at scholars in the humanities: (a) The Prehistoric Archaeology of the Aegean [32], (b) Paedeia, a site on ancient literacy and (c) the Hear Homer site, both under construction [7], as well as the ASML pages themselves [25].

There are many approaches to the creation of resources for the World Wide Web. Documents are electronically served in many forms including raw text, hypertext markup language (HTML), graphic images, and executable programs. The source mechanism for content is either transmission of static files or dynamic generation of content from programs (often called scripts). By far the most common mechanism is the delivery of static files in HTML format with enclosures (in-line components) usually in the form of images.

Web sites provide detailed multimedia presentations and are highly diverse in interactivity, visual content richness and overall design. These presentations must be created (authored) using any one of several approaches. Often content is created by simply writing HTML manually using a text editor. Products also exist which allow for authoring pages without knowledge of HTML in a form similar to working with a word processor. The typical authoring process of a web site is the “page at a time” approach which views a site as a collection of discrete hypertext documents. DEVLAB experiences with web site
development projects, such as the *DAGS’95 Conference Proceedings* [8], *The Olympics in the Ancient Hellenic World* [1], and *The Prehistoric Archaeology of the Aegean* [32] have clearly exhibited the shortcomings of the page-at-a-time approach. Rice et al., discuss similar experiences creating large WWW sites, wherein they state that “trying to correct many of our early errors has become an extremely time-intensive job” [29]. They specify a design process that includes defining structure and format prior to page creation, but clearly found that such decisions cannot be expected to be perfect and later revision can require editing every page in a site. Creating the pages of a site one at a time is similar to creating a word processing document one page at a time. Were this paper created this way, the authors would be forced to edit every single file wherever a change was made to overall document format.

One common solution to this problem has been to “build” sites using custom scripts in computer languages such as Perl [32]. This cumbersome approach has been used in the DEVLAB to create smaller sites for internal use. Other automated solutions also exist; for example, the web site for a series of Dartmouth conference proceedings, the DAGS’95 proceedings, was built using an Apple Hypercard stack [8]. The Olympics in the Ancient Hellenic World site was created using scripts based on m4, a macro language common in UNIX systems, and Perl scripts for dynamic content generation [1]. Many other projects in the DEVLAB have been done using the Perl language entirely. These projects assumed a knowledgeable programmer with experience in programming with the Perl language and the need for *custom software development for every single project*. DEVLAB projects often involve novice programmers as well as students not associated with Computer Science. Indeed, there are two aspects associated with web site development which motivated the development of ASML: (a) most web sites are built with the assistance of novices who do the content processing and collection and (b) most web sites require extensive maintenance and update which is hard to do for each page. Both of these features suffer by the one-page-at-a-time approach. ASML was developed to better support authoring and maintenance in a fast-changing environment.

ASML incorporates many features which automate the development of sites. It can be thought of as a “markup language for markup languages.” Figure 1 illustrates the use of

![Figure 1](image)

*Figure 1.* Use of ASML at a site.