

PANEL: HTML: POISON OR PANACEA?

Moderator: Robert Glushko, Passage Systems (USA)

Panel:

Dale Dougherty, O'Reilly & Associates (USA)

Eliot Kimber, Passage Systems (USA)

Antoine Rizk, Euroclid (France)

Daniel Russell, Apple Computer (USA)

Kent Summers, Electronic Book Technologies (USA)

Many people are having their first experience with a distributed hypertext system by using Mosaic or some other viewer based on HTML, the HyperText Markup Language of the World Wide Web. HTML's simplicity allows it to be created without special authoring tools or expertise, and the ubiquity of free WWW viewers like Mosaic removes one of the cost barriers. Because HTML is an application of SGML, the Standard Generalized Markup Language, it has also introduced many people to the concepts and syntax of application-independent markup. The explosive growth of the WWW makes it undeniable that HTML and Mosaic will serve as the reference point for much future thinking about hypertext and SGML outside of the academic and research community.

But if HTML is to some people a democratizing force for hypertext authoring and publishing, to others its lack of structure and validation is a substantial step backwards for authors, readers, and their organizations. HTML has proven the basic premise of SGML, that a standard, application-independent data representation can enable blind interchange among disparate and even unknown or unpredicted applications. But to many SGML advocates, HTML is too primitive, and today's HTML documents will end up as tomorrow's cast-off legacy data format. HTML viewers may seem appealing for network publishing, but to some they are merely seductive and superficial, undermining years of careful research on usability and business models.

The goal of this panel is to de-balkanize the bi-polar HTML camps and seek a productive role for all points of view. HTML and the WWW are simply too important for the academic and research community to ignore. HTML will not go away. We must participate in the evolution of HTML and find ways to become part of the network publishing revolution it represents.

Dale Dougherty, O'Reilly & Associates:

HTML is the best thing that has ever happened to SGML. The spectrum of choices for online delivery formats has SGML at one end and PostScript at the other. HTML is helping to concentrate attention at the SGML end and it has hurt proprietary formats like Acrobat.

The greatest strength of HTML is that it defined a globally resolvable link addressing mechanism. These addresses, called Universal Resource Locators (URL) aren't perfect, either, but the approach is simple, and allows you to do amazing things. Like HTML itself, URLs will be refined over time but the key concepts have been realized today.

HTML has succeeded because it started small, and indeed it is only a starting point, not a final destination. The success of HTML can be attributed not to the design of the language, but HTML exists as a public specification and that software is widely available to render HTML. Standards wonks take note: providing free software that implements an open, public specification is the new road to developing standards.

Eliot Kimber, Passage Systems:

HTML proves the basic premise of SGML the ability to interchange data among applications using a vendor-neutral data representation. But HTML fails to embrace the larger vision of SGML and ignores many of its technical details, in the process duplicating with application-specific constructs and constructs already defined by the SGML and HyTime standards.

In addition, the lack of true SGML processing provided by HTML browsers has resulted in data chaos in spite of the power of SGML to bring order to documents. In many ways, HTML is no better than any other proprietary word processing format because what is perceived to be "good HTML" is what looks good under Mosaic, not what conforms to the HTML document type.

Finally, the existence of HTML as a document type, and not a document architecture, shows a lack of understanding of the true purpose of SGML, which is to serve data owners, not tool providers, and that imposing any data format, even an SGML data format, denies data owners their right to define the structure of their own data, and does so needlessly. In this sense, HTML is no better than Adobe Acrobat, offering only a cheap way to get data on a screen, but doing nothing to further the needs of data owners to more efficiently control and use their own data.

Antoine Rizk, Euroclid:

How could we be satisfied with HTML when we believe that even SGML is not sufficient and had to develop HyTime? Internet publishing problems could not be resolved by simply providing an exchange format that is the least common denominator of all available platforms. The problem with Wide Area Network publishing is the gap that exists between content providers and final readers. What is needed is tools for hypermedia editing that would act as mediators and assemblers of different sources of information.

We have addressed these issues and implemented in conjunction with France Telecom a WAN link server that uses MHEG as an exchange standard. Content providers 'rent' link server space to represent their services, hypermedia application providers use the link server to interconnect heterogeneous data from different sources, and readers login via a Kiosk service and navigate.

A single format will never suffice, no matter how rich it is. We need to provide means for accessing and interconnecting heterogeneous, standard and proprietary formats.

Daniel Russell, Apple Computer:

The beauty of HTML is that it's incredibly simple. As served up by the WorldWideWeb, HTML documents (while admittedly not things of beauty or sophistication) are tremendously straight-forward to create. As a consequence, without any of the overhead of the full SGML mindset, HTML documents have multiplied prolifically. Frequently, the only instruction needed to bring authors up to speed are a few examples of raw, seething HTML text and the image of the page when rendered. While this simplicity is potentially damaging to the existing SGML tutorial business, it has admitted a huge number of participants to the global hypermedia environment through WWW.

Like HTML, the entire Web works because it is simple. The protocols are straight-forward and easy to implement. The system design is not overburdened with creeping feature-osis, and there isn't (yet) a successful effort to corral the enthusiasm of Web authors behind the veil of bureaucracy.

The challenge ahead will be to engineer new additions to the HTML & the WWW so that the simplicity remains while power and usefulness to its users improves.

Kent Summers, Electronic Book Technologies:

HTML is a double-edged sword. Because of its simplicity, HTML doesn't require any special authoring tools or expertise, which is one important reason why the Web is so popular. However, it is widely agreed that HTML is ineffective for representing large or highly structured documents. The current Internet environment has no acceptable solution for handling complex documents which require a level of structure and formatting much higher than can be achieved with HTML. In addition, retrieving and viewing large documents-anywhere over the 250K range-is unbearably slow for the majority of users.

Electronic Book Technologies develops the DynaText publishing system that is used throughout the world for SGML-based publishing. Our strategy is not to deny the existence of HTML and the ubiquitous low-function viewers it has spawned, since many of our customers are also interested in HTML. Instead, our strategy with our new DynaWeb system is to bring the best of these technologies together, opening up the Internet to industrial publishing opportunities for the first time.