Editor: Ron Vetter, North Dakota State University, IACC Bldg., Rm. 258, Fargo, ND 58105-5164; phone, (701) 231-7084; Internet, rvetter@plains.nodak.edu

Using the WWW Test Pattern to check HTML client compliance

Hal Berghel
University of Arkansas

Certainly the hottest part of the Internet right now is the World Wide Web. Since its inception in 1990, the Web has proven to be the unifying environment for the digital resources of the Internet. By all measures, it is enormously successful. Consider that in just a few years the Web has come to be the leading Internet resource, providing 21.4 percent of the total packet count and 26.3 percent of the total byte count on the NSF backbone. This compares with 14.0 percent and 21.5 percent for ftp, 8.1 percent and 8.6 percent for nntp (Network News Transfer Protocol), 7.5 percent and 2.5 percent for Telnet, and 1.5 percent and 1.8 percent for Gopher.[1] This is even more remarkable given that the Web really didn't take off until 1992 when the first navigator/browsers became available. There is no longer any question that the World Wide Web has evolved into an indispensable resource for the networking community.

But it's not yet a perfect cyberworld. A major difficulty lies in the inconsistency with which Web client developers comply with the emerging standards, and this inconsistency translates into headaches for end users. The University of Arkansas Web Resources Group has set out to alleviate this problem. We have created a cybermedia test pattern for the Web; below we describe its purpose and use.

The Web's protocols

As with other Internet services, the nuts-and-bolts part of the Web is a set of client-server protocols. The first, Hypertext Transfer Protocol (HTTP), provides a uniform handshaking and format protocol for client-server communication. The client establishes a connection with the server, makes a request, receives a response, closes the connection, and takes action. In the simplest of cases, a set of files of varied media are requested from the server to be displayed by the client-side navigator/browser.

The second protocol, Hypertext Markup Language (HTML),[2] defines the internal structure of the Web's "documents." It accomplishes this through a primitive tagging convention that identifies contained or referenced resources. For example, a sensitive (clickable) document anchor that points to a uniform resource locator (URL) would be couched within the tag pairs <A HREF = ...> and </A>; an image would be identified by the tag <IMG SRC = ". . . .">, and so forth. Though unsophisticated, it works—at least for the most part.

The HTML protocol has evolved in stages, or levels, over the past three years, and this evolution has precipitated some discomfort. The compliance levels are specified by the World Wide Web Consortium,[3] but developers do not follow the prescribed specifications consistently enough.

HTML level 0 provided specifications for basic HTML structure. Level 0 had support for hypertext links. However, it had only meager format control and limited text enhancements. Level 1 defined extensions for basic image handling, limited text enhancement, and relative resource addressing. Level 2 included specifications for forms, along with incremental gains in the other areas defined for levels 0 and 1. Level 3 will provide extensions for tables, a LaTex-like ASCII-notation standard for mathematical formulas, and features for additional multimedia support. That comes to four compliance levels in just under three years.

To make things worse, Web client conformance is usually discussed in the context of HTML versions. The HTML Version 1 convention includes levels 0 and 1 standards. HTML Versions 2 and 3 include levels 0-2 and 0-3, respectively. However, the HTML version numbers are really only discussed in the abstract, because typical Web-client developers make no claims of compatibility—they usually add as many features as they feel they can manage before a new release and let it go at that. Even if users understood what was involved in these compliance issues, there would be no way to relate it to a particular product.

But it doesn't end there. Nonstandard extensions are also emerging in parallel with orthodox versions. This,

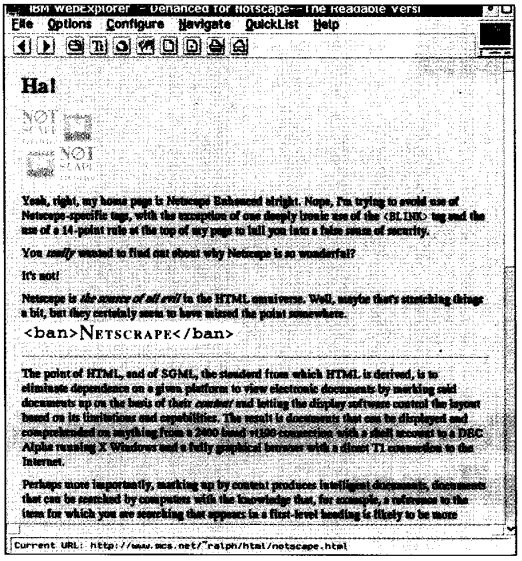

Figure 1. Anti-Netscape crusade. This particular navigator/browser, Web Explorer, does not support many of the Netscape extensions. If it did, this page would be virtually unreadable—which is the author's intention. The effect is most pronounced when viewing the page with side-by-side navigator/browsers (URL=http://www.mcs.net/~ralph/html/notscape.html).

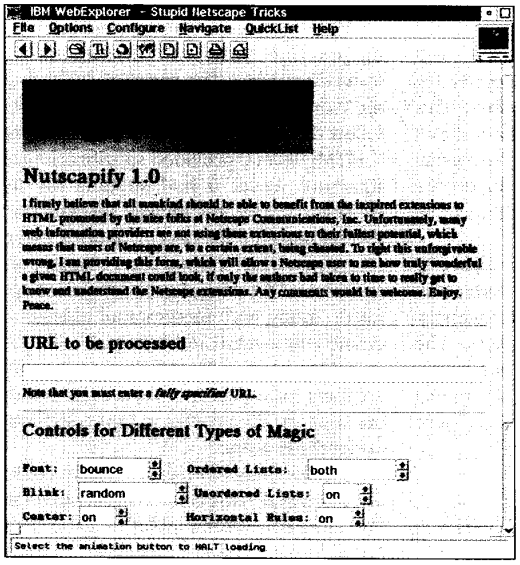

Figure 2. An imaginative attempt to highlight the potential of Netscape extensions. Be forewarned that non-Netscape clients may behave strangely (URL=http://thule.mt.cs.cmu.edu:8001/tools/nutscape/).

together with the sometimes conflicting interests of commercial versus not-for-profit developers, is the battlefield of a technology skirmish (see Berghel[4]). In general, the nonstandard extensions apply to the body of HTML documents and are associated with a particular Web client, Netscape. Extensions dealing with image alignment and resizing, box graphics, and greater control over type size and font are commonly used Netscape extensions.

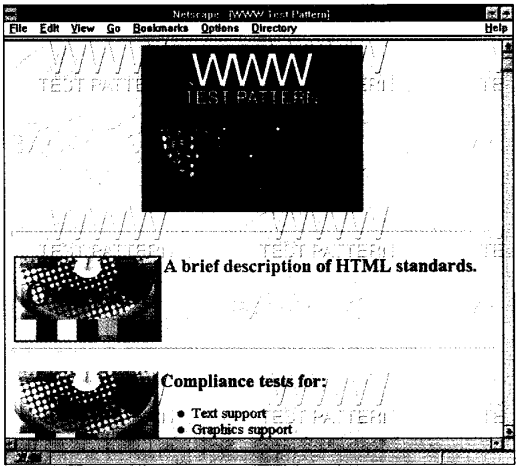

Figure 3. The tiled background to the home page is rendered correctly by Netscape Version 1.2.b2.

We will ignore for the moment the problems of feature imbalance for the same product across multiple platforms, as well as the problem of implementation bugs; these difficulties relate to the lack of client navigator/browser uniformity. (For further details, see our forthcoming "The Client Side of the Web."[5])

HTML compliance: Evolution or revolution?

So quite apart from an orderly evolution, the current state of HTML compliance also suggests a degree of *revolution*. This explains most of the current discomfort on the user's side of the Web. To the user, this lack of uniformity surfaces in improperly rendered media, incorrect display formatting, forms that aren't seamlessly linked to their PERL (Practical Extraction Report Language) scripts, and so forth.

To illustrate the scope of the problem, of the eight primary navigator/browser clients used in our lab, only two fully comply with all HTML level 0 specifications. While the occasional deficiencies (for example, the rendering of menu, directory, and unordered list element tags) are not earthshaking, they can be irritating. This problem gets worse as we escalate HTML levels, until we reach a free-for-all at level 3. Enter into the mix the fairly widespread acceptance of a few Netscape extensions, and the result is some real confusion over standards and some hard-to-read Web documents.

This conflict over standards has even become politicized on the Net. At this writing there are actually "digital campaigns" for and against Netscape extensions (see Figures 1 and 2). While little of any enduring value will likely follow from this activity, its very occurrence suggests the presence of some important underlying issues.

The World Wide Web Test Pattern

The HTML compliance issues will not be resolved anytime soon; anarchy is always hard to orchestrate. Web

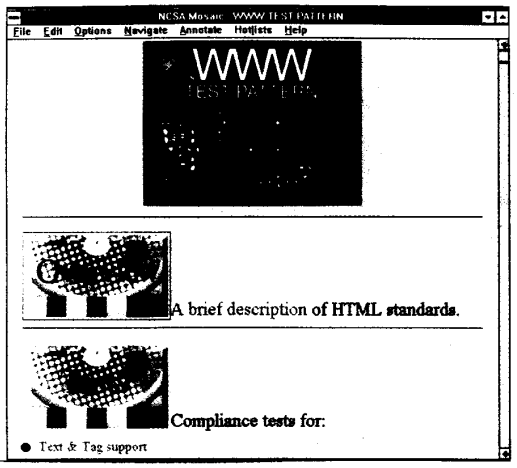

Figure 4. The background from Figure 3 is not rendered at all by NCSA Mosaic Version 2.0.0b4.

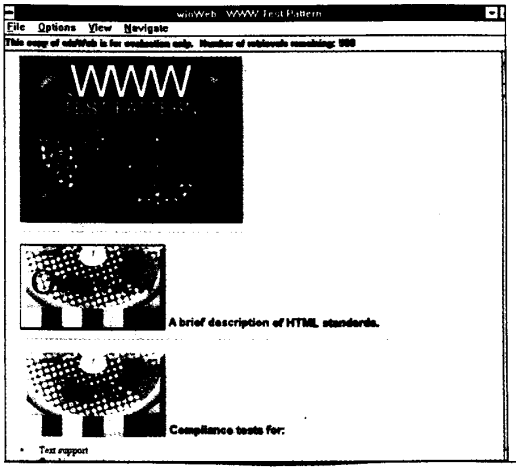

Figure 5. This version of Figure 3 shows yet another variation in the degree of compliance.

clients will come and go. Within a few years, the descendants of those that survive will eventually be bundled with operating systems or Internet connectivity packages, or be seamlessly integrated into desktop suites. Perhaps by then we will have de facto if not de jure standards in place. But for now we have information to process even while many of our Web resources are presented to us in disarray.

Enter the World Wide Web Test Pattern. This Web site was conceived as a general-purpose test bench so that users and developers can check for HTML compliance. While still under construction, it already includes a standard suite of tests for text, audio, graphics, meta-links, animations, forms, and tables. The URL is http://www.uark.edu/~wrg/.

Figures 3 and 4 illustrate how the Web Test Pattern may be used. Observe that there is a tiled background to the WWW Test Pattern home page, which is rendered correctly by Netscape Version 1.2.b2 (Figure 3) but not rendered at all by NCSA Mosaic Version 2.0.0b4 (Figure 4). Tiled background is an element of the proposed HTML 3 specifications.

The subtle change in Figure 5 indicates that there can be gradations of compliance. In this case, not only is the background missing, but also the superimposed image is not properly centered. Clearly, winWeb 1.1 B1.2 is not up to the challenge.

Some of the tests, like those in Figures 3-5, are passive; the user merely loads the test document and views the result. Other tests require direct user involvement. Audio files provide a case in point because they are never in line, even though their players may be integrated into the client. That is, unlike in-line image files, which are automatically displayed, audio files require an action on the user's part to be played—either with an internal player built into the client or with a spawnable external player. Most modern clients include user-configurable launch pads, so over time the importance of the distinction between integrated and spawnable perusers will vanish.

Currently, cybermedia tests exist for the Netscape "server-push" and "client-pull" extensions as well as for MPEG, AVI, and QuickTime animations. We hope that the entire HTML level 3 suite will be operational by the time this article appears in print.

As it develops, the Web Test Pattern will attempt to include as rich a variety of media as there is on the Web, thereby enabling both users and developers to test for compliance with HTML levels.

THE WEB TEST PATTERN IS AVAILABLE for use by both Web users and developers for monitoring the degree of HTML compliance of Web clients. We are investigating the viability of reducing the multiplicity of tests and providing a standardized report.

Acknowledgments

The World Wide Web Test Pattern was produced by the University of Arkansas Web Resources Group, which includes Jon Ashley, Troy Cash, Peter Laws, John Wiggins, and myself. We thank Ron Vetter for encouraging us to write this article for *Computer*.

References

1. *NSFnet Backbone Traffic Distribution Statistics*, Apr. 1995, http://www.cc.gatech.edu/gvu/stats/NSF/merit.html.
2. World Wide Web Consortium, *Hypertext Markup Language*, http://www.w3.org/hypertext/WWW/MarkUp/MarkUp.html.
3. *World Wide Web Consortium*, http://w3.org/.
4. H. Berghel, "OS/2, UNIX, Windows, and the Mosaic Wars," *OS/2 Magazine*, May 1995, pp. 26-35.
5. H. Berghel, "The Client Side of the Web," *Comm. ACM*, Jan. 1996 (forthcoming).

Hal Berghel *is professor of computer science at the University of Arkansas and chair of the Board of Directors of the Arkansas Society for Computer and Information Technology. His Web address is http://acm.org/~hlb/homepage.html; e-mail hlb@acm.org.*