Reflections on "Seven Issues": Hypertext in the Era of the Web

Frank G. Halasz
130 Bradley Dr.
Santa Cruz, CA 95060
fghalasz@pacbell.net

Abstract

Frank Halasz revisits his papers "Reflections on NoteCards: Seven Issues for the Next Generation of Hypermedia Systems" and "Seven Issues: Revisited" from a Web perspective. His goal is to better understand how the Web has shaped and often overshadowed how we look at the tremendous possibilities of hypertext.

Keywords

hypertext/hypermedia, user interface design, World-Wide Web

Introduction

In the thirteen years since the publication of "Reflections on NoteCards: Seven Issues for the Next Generation of Hypermedia Systems" (Halasz, 1988), the world of hypertext has been swept by a tidal wave of change brought on by theWorld-Wide Web. Culturally, economically, and practically, the Web has come to dominate the world of hypertext, as it has for much of the world of computing in general. Even in research, the Web dominates much of the thinking about advanced hypertext systems and applications. Looking at it with the benefit of hindsight, I see the most remarkable aspect of "Seven Issues" is that it missed the Web entirely.

In writing "Seven Issues," my goal was to lay out a vision for the near-term future of hypertext, circa 1988. Three years later, in my "Seven Issues: Revisited" talk at Hypertext'91 (Halasz, 1991), my goal was to review and renew that vision based on the progress in hypertext from 1988 through 1991. In rereading "Seven Issues" and "Revisited" for the first time in almost 10 years, I was struck by the fact that neither of these visions included even a hint of the world-wide, world-transforming hypertext system that was just around the corner. As pointed out by Elli Mylonas in her commentary in this issue, "Seven Issues" is about systems "in the small." From the "Seven Issues" perspective, hypertext is a tool for individuals and workgroups to manage their unstructured information. Hypertexts in the large—world-wide repositories of interconnected information nodes —just don't factor into the thinking in "Seven Issues."

ACM Journal of Computer Documentation, 2001,25:109-114
 ISSN 1527-6805

The notion of hypertexts in the large had been around long before "Seven Issues." In fact, the term *hypertext* was coined by Ted Nelson in the 1960s while describing just such an all-encompassing information web (for example, see Nelson, 1987). But Nelson's grand vision was not implementable, and at the time I (and I believe most of my fellow hypertext researchers) felt that the sensible thing to do was to focus on much smaller, more task-oriented, implementable hypertexts. This bias was so strong that I actually missed the first glimmerings of the Web. I recall reading an early version of one of the first papers on the World-Wide Web by Tim Berners-Lee and colleagues (see Berners-Lee et al., 1994) a few months before the Hypertext '91 conference. I largely dismissed the paper as yet another impractical Nelsonian vision. As a result, the concept of HTML and the Web made little impact on the issues I raised in "Revisited."

Despite my failure to foresee the Web revolution in "Seven Issues" and "Revisited," much of the thinking in these two papers is still relevant in the Web era. In this commentary I revisit the papers from a Web perspective. My goal is to better understand how the Web has shaped and often overshadowed how we look at the tremendous possibilities of hypertext.

Taken together, "Seven Issues" and "Revisited" present a total of seven dimensions that can be used to characterize the space of hypertext systems as well as a total of thirteen issues that in 1988 and 1991 defined my vision for the future of such systems. In the remainder of this commentary, I consider in turn each of these dimensions and issues from a Web perspective, aggregating where necessary to account for considerable overlap in ideas amongst the various dimensions and issues.

Dimensions of the Hypertext Space

Dimension 1: Scope

This dimension was intended to highlight the fact that different design points and issues arise in building hypertext systems as they range in scope from those intended to be used by a single person running on a single workstation (for example, NoteCards; Halasz et al., 1987) to those intended to be used to store all of the world's literature (for example, Xanadu; Nelson, 1987). Fundamentally, the Web is at the latter end of this spectrum. It is a world-wide distributed hypertext containing billions of nodes and links. However, one of the most remarkable aspects of the standards and technologies that underlie the Web is their tremendous scalability. The same mechanisms that power the enormous public Web are also used to power collaborative document management systems used by small workgroups. They also power the help system for many desktop applications and even run, in somewhat restricted form, on cellular phones. If I were to rebuild NoteCards today, I would implement it using Web technologies. In this sense, the dimension of scope has less relevance in the era of the Web.

Dimension 2: Browsing versus Authoring

This dimension distinguishes systems in which the primary user is a consumer of "published" hypertexts from systems in which the primary user functions as both a producer and a consumer of the hypertext. Fundamentally, the Web belongs in the former category. Its dominant usage pattern is one in which a few publish Web pages that many browse. There are, however, a growing number of websites and Web-based systems, primarily found on corporate intranets, that are designed to support authoring. Collaborative document management systems such as Xerox DocuShare (docushare.xerox.com) are an example. The WebDAV suite of protocols (see www.webdav.org) is an attempt to standardize some of the mechanisms that underlie such authoring-based websites.

Dimension 3:TargetTask Domain

In the 1980s, most hypertext systems were designed with a specific task domain in mind. Although the builders of these systems were generally attempting to build general hypertext facilities, their architectural and implementation decisions were heavily influenced by the target task domain. This is less true of the Web, which hosts an amazing variety of websites supported by a very general set of mechanisms, including TCP/IP, HTTP, HTML, and Javascript. In addition to these general mechanisms, however, the Web incorporates a crazy amalgam of overlapping and often very task specific technologies and standards. A quick look at the World Wide Web Consortium website (www.w3c.org) reveals the following standards activities: SMIL (Syn-

chronized Multimedia Integration Language), Common Micropayment Markup, MathML, and TVWeb. Each of these activities is an attempt to create a standardized mechanism to support some specific task domain within the greater set of Web technologies. No one can deny that the Web is a truly general hypertext, but its generality depends to a great degree on these task-specific standards.

Dimension 4: The Navigators versus the Architects

In "Revisited" I made the distinction between the Navigators and the Architects. The Navigators are hypertext systems in which the focus is on the nodes and their contents. Links are just a means to get from one node another. The Architects are hypertext systems where the main focus is on the structure of the network. Links, and to some extent nodes, are primarily a means of creating this structure. "Seven Issues" was written with a strong Architect bias. It views hypertext as a knowledge representation tool as well as an information management tool. Links are much more than a navigation mechanism. They are part of a network structure that is used to represent the information contained in the hypertext. The graphical network browsers found in NoteCards are a manifestation of this view of hypertext as information representation.

In contrast, the Web is very much a Navigator. It is essentially a node-centric hypertext where the web page dominates and the notion of information representation through linked structures is largely absent. I have never encountered a node-link browser of the kind found in NoteCards on any website. (Given the vastness and diversity of the Web, however, I can't say with absolute certainty that one doesn't exist.)

Dimension 5: The Card Sharks versus the Holy Scrollers

In "Revisited" I adopted this wonderfully named dimension from Jef Raskin's Hypertext '87 talk (Raskin, 1987). To a Card Shark, each node in the hypertext is a fixed-size "card" onto which attached items (text, graphics, animations, etc.) can be placed. Navigation consists of jumping from card to card. To a Holy Scroller, the hypertext is a collection of lengthy documents and navigation consists of "jumping" within and between documents as well as scrolling within a document. In 1991, this distinction provided some insight into the underlying architecture and implementation of various hypertext systems. In the Web era, this distinction seems vacuous. Both styles of web pages, as well as many in between, can be found on the Web, often intermixed within a single site. The distinction is more an issue of style and page design than a fundamental difference in hypertext architecture.

Dimension 6: The Literati versus the Engineers

In using this rather fuzzy distinction, I was attempting to distinguish between hypertexts intended to be a new form of literary expression (the Literati) and hypertexts intended to merely to convey some structured information (the Engineers). The distinction was intended to acknowledge that by 1991 the hypertext research community had two rather different facets: the engineers building hypertext systems and the literary scholars and writers of hypertext fiction exploring a new medium of expression. That distinction still holds today. And although the Engineers dominate the headlines when it comes to the Web, there is a growing body of true hypertext literature distributed on the Web and otherwise (see www.eastgate.com). In hindsight, however, it is clear that I missed perhaps the most important players, who in 1991 were still waiting in the wings: the Business People. In the world of the Web, the Literati are but a drop in the bucket compared to the Business People who see in hypertext a whole new medium for buying, selling, advertising, and distributing goods and services.

Issues for the Next Generation of Hypertext Systems

Issue 1: Search. Even in relatively small hypertext networks, finding a specific piece of information solely by following links is an arduous task. In NoteCards, we found that it was very easy to "lose" information in networks with as few as 500 cards. Based on these experiences, "Seven Issues" highlights the importance of using automated search to complement (or even replace) link traversal in finding information in a hypertext. Given that the Web is many orders of magnitude larger than 500 nodes, it is no surprise that automated search has become one of the cornerstones of Web access. Search engines are practically the only way in which a user can efficiently locate specific information in the vast sea of pages on the public Web.

"Seven Issues" draws a strong distinction between content search and structure search. In content search,

each node and link is examined independently to determine if it contains the queried information. In structure search, the hypertext network is searched for a specifically shaped subnetwork of nodes and links. In the node-centric world of the Web, content search predominates. Some search engines do use a bit of the Web's interconnection structure to aid in content-based search. Google (www.google.com), for example, uses the number of links referring to each web page to help weight the results of a query (Brin and Page, 1998). But a search engine that looks for structural patterns makes little sense given the lack of meaningful node-link structure inherent in the Web.

Issue 2: Composites. One of the significant weaknesses in many second-generation hypertext systems was the lack of a facility to represent and manipulate groups of nodes and links as unique entities separate from their components. "Seven Issues" proposes that this weakness can be overcome by adding composition as a fundamental construct to the hypertext data model. The Web has progressed through a similar evolution. In its initial stages, the Web consisted primarily of individual pages connected by links. In 1998, the release of HTML 4.0 provided several new mechanisms for creating composite web pages including Frames and embedded HTML Objects (see www.w3.org/TR/1998/REC-html40-19980424). Although these mechanisms embody a somewhat weaker notion of composition than is described in "Seven Issues," they are well suited to the node-centric, "publish/browse" nature of the Web. Today a very large percentage of Web pages use these mechanisms to combine two or more web pages into a single page for presentation to the user.

Issues 3 and 4: Virtual Structures and Computation In/Over the Hypertext. (Also Dimension 7: The Literalists versus the Virtualists). Many hypertext systems are limited to passive storage and retrieval of information. These systems have trouble dealing with rapidly changing information. "Seven Issues" 0proposes that these systems could be made more adaptive by the addition of virtual structures, that is, nodes, links, and composites whose contents are computed each time they are accessed, and by the use of computational engines that "crawl" the hypertext, adding, deleting, and modifying its structure and content.

Virtuality is rampant on the Web. Any Web page you look at is as likely to have been computed on the fly as it is to have been pre-authored in HTML and stored for your retrieval. This is true for both "singleton" Web pages and composite Web pages. Surprisingly it's also true for many links. The Web-based bank statement that links to Web pages containing images of your cleared checks has in all likelihood been created at access time using a query on an external accounts database.

Computational engines are also commonplace on the Web. The major search engines are continually crawling the Web, indexing each Web page they find. Similarly, comparison shopping sites such as MySimon (www.MySimon.com) are continually crawling the contents of other online shopping sites in order to compile a directory of available goods and their prices. Although these computational engines differ in nature from the AI-based inference engines highlighted in "Seven Issues," they serve the same purpose of keeping the hypertext up to date in the face of rapidly changing information.

Issue 5: Versioning. Versioning was one of my original seven issues. However, I dropped it in "Revisited" because of a complete lack of interest in the topic within the hypertext community. Versioning is also essentially nonexistent on the public Web, where there are no standard mechanisms for versioning Web pages and links. Many individual Web sites and intranets, however, do implement proprietary versioning mechanisms in support of specific applications such as software development and collaborative document management. As mentioned above, the WebDAV group has proposed extensions to HTTP to standardize the use of versioning on the Web for such applications.

Issue 6: Collaboration In/Over Hypertexts. "Seven Issues" divides support for collaboration into two somewhat distinct subissues: 1) infrastructure for supporting collaboration and 2) support for the social interactions involved in the use of a shared network. Transaction support, concurrency control, and change notification are examples of the infrastructure needed to support collaboration. Given the preponderance of "publish/browse" usage on the public Web, infrastructure for supporting collaboration is generally not necessary. One interesting exception is change notification, which has become a standard feature on many public Web sites. These sites will, for example, send an e-mail to a user when a specific Web page or element on a Web page has changed. Also, as in the case of versioning, a number of task-specific Web sites and intranets

have implemented proprietary mechanisms to support collaboration. Here too, WebDAV is an attempt to standardize some of these mechanisms.

In contrast to the need for a collaborative infrastructure, the issue of supporting the social interactions involved in using a shared network touches each and every Web user. This issue is essentially one of usability and "readability" of the Web. The rapid growth and commercialization of the Web has resulted in a wide variety of methods and styles for authoring Web pages, some more intelligible than others. Interest in improving the overall usability of the Web is high. Attempts to look at this issue span from literary criticism and rhetorics (see, for example, Landow, 1992) through classical usability analysis (see, for example, Nielson, 2000) to commercial usability assessments (see, for example, www.vividence.com). The goal of all these efforts is to make information and interactions on the Web more accessible, whether for collaborative authors or for casual Web browsers.

Issue 7: Extensibility and Tailorability. "Seven Issues" argues that hypertext systems should be easily extended and tailored to meet the needs of specific tasks and users. The Web is just such a hypertext system. With a basis in open standards and backed by a cornucopia of both client and server-side interfaces to computation and scripting engines, the Web is almost infinitely extendable and tailorable to meet the needs of any user or task. Moreover, many of these wonderful facilities (HTML and Javascript, for example) are easy enough to learn and use that they meet the challenge that I put forward in "Seven Issues" of enabling "non-programmers" to build interesting hypertext applications.

Issues 8 and 9: Open Systems and Standards. In "Revisited," I added Open Systems and Standards to the original list of seven issues. My belief was that the self-contained, monolithic hypertext systems that were designed in the 1980s would be replaced by open systems in which independent but communicating components would work together to produce the hypertext functionality. Such open systems could exist only if we developed and adopted standards by which the various components could communicate and exchange data. This vision seems tame compared to what has actually developed. The Web is in fact one giant open hypertext system in which an enormous assembly of interconnected components (desktop browsers, web servers, search engines, etc.) communicate and exchange data by using an amazing array of standards, including communications protocols (HTTP, HTTPS, FTP, etc.), document formats (HTML, JPG, etc.), scripting languages (Javascript, Perl, Tcl, etc.), and APIs (CGI, etc.). Open architectures and standards are the two key principles that made it possible for the Web to grow and develop at the phenomenal rate that it has over the last 10 years.

Issue 10: User Interfaces for Large Information Spaces. In "Revisited," I predicted that novel interfaces that allow users to manipulate large network structures on a workstation screen would become a critical factor as larger hypertexts became more prevalent. With the advent of the Web, larger hypertexts have certainly become more prevalent. But the Web's node-centric hypertext model simply doesn't require users to deal with large network structures. Novel user interface technologies such as fish-eye views (Furnas, 1986) and treemaps (Johnson and Shneiderman, 1991) that help users deal with large network structures are for the most part unnecessary. Instead, the vastness of the Web demands technologies that allow users to find their way directly to the information/entertainment that they seeking. Portal sites and search engines are the tools of choice here.

Issue 11: Very Large Hypertexts. In "Revisited," I challenged the hypertext community to build usable industrial-strength hypertext systems capable of handling 10,000 or even 100,000 documents. A hundred thousand documents, can you imagine that? In hindsight, this challenge seems positively quaint. The last time I looked at Google (www.google.com), it had indexed over 1.3 billion Web pages. The scale of the public Web exceeds by several orders of magnitude anything contemplated in "Seven Issues" and "Revisited."

Issues 12 and 13: Hypertext Markets and Publishing Hypertexts. In 1991, it still wasn't clear how hypertext was going to make the leap from experiments and prototypes into commercially successful products and projects. "Revisited" encouraged the hypertext community to focus less on pure technology and pay more attention to the factors necessary to bring hypertext to the commercial market, including the problem of bootstrapping a hypertext publishing industry. In retrospect, this encouragement was completely unnecessary. The Web has shown that there is a huge (although as yet questionably profitable) market for hypertexts and hypertext applications. And although it would be hard to pinpoint a substantial hypertext publishing industry of the sort envisioned in "Revisited," there are billions

of hypertext pages already on the public Web, with many thousands more being published every day.

It is interesting to note that the market has never really developed for the kind of representationally oriented "hypertexts in the small" that formed the background to "Seven Issues" and "Revisited." Individual and small-group hypertext systems and hypertexts that focus on the network structure *per se* are still largely confined to research laboratories.

Conclusion

As Rick Furuta points out in his commentary in this issue, the popularity of the World-Wide Web caught the hypertext community by surprise in part because the hypertext model embedded in the Web is one that the community had long evolved past. But, counters Cathy Marshall in her commentary, would the Web have taken off if it were as complicated as the hypertext systems and models envisioned in "Seven Issues"? Clearly not. The power of the Web lies to a great degree in its simplicity and, to borrow Furuta's phrase, its intensely pragmatic architecture. It is this simplicity and pragmatic approach that have allowed the Web to prosper and grow beyond anything imagined in "Seven Issues."

The Web's straightforward and practical approach, however, does not negate the fact that it is inherently a hypertext "system." It should be no surprise that today's Web designers and implementers face many of the same design choices and problems that we encountered in building the second-generation hypertext systems during the 1980's. I would hope that the dimensions and issues raised in "Seven Issues" and "Revisited" are as insightful for these Web designers as they were for those of us in the pre-Web hypertext community.

In closing, I feel compelled to note that even though I firmly believe that "Seven Issues" remains relevant in the Web era, it is not a paper I would republish under the title "Seven Issues for the Web." The practicality, popularity, and scale of the Web have brought to the fore many issues that I never even considered while writing "Seven Issues" and "Revisited." Many of these issues, including security, privacy, and commercial transactions, are very significant and would clearly take precedence over any of my "seven" issues in the list of seven issues for the next generation of the Web.

References

Berners-Lee, T., et al. (1994). The world-wide web. In *Communications of the ACM,* 37 (8), 76-82.

Brin, S. and Page, L. (1998). The anatomy of a large-scale hypertextual search engine. In *Proceedings of the 7th International World Wide Web Conference* (Brisbane, Australia, April 14-18), 107-117.

Furnas, G.W. (1986). Generalized fisheye views. In *Proceedings of the 1986 ACM Conference on Human Factors in Computing Systems (CHI '86)* (Boston, Mass., April 13-17), pp. 16-23.

Halasz, F.G. (1988). Reflections on NoteCards: Seven issues for the next generation of hypermedia systems. In *Communications of the ACM,* 31 (7), 836-852.

Halasz, F.G. "Seven issues": Revisited (1991). Keynote talk at the 1991 ACM Conference on Hypertext (Hypertext '91) (San Antonio, Texas, December 18). Slides available as Technical Report #SSL-91-149 from Xerox Palo Alto Research Center, Palo Alto, Ca. Transcript available at http://www.parc.xerox.com/spl/projects/halasz-keynote.

Halasz, F.G., et al. (1987). NoteCards in a nutshell. In *Proceedings of the 19Conference on Human Factors in Computing Systems (CHI+GI '87)* (Toronto, Canada, April), 345-365.

Johnson, B. and Shneiderman, B. (1991). Treemaps: a space-filling approach to the visualization of hierarchical information structures. In *Proceedings of the 2nd International IEEE Visualization Conference* (San Diego, Oct.), 284-291.

Landow, G.P. (1992). *Hypertext: The Convergence of Contemporary Critical Theory and Technology.* Baltimore: The Johns Hopkins University Press.

Nielsen, J. (2000). *Designing Web Usability: The Practice of Simplicity.* Indianapolis: New Riders Publishing.

Nelson, T. (1987). *Computer Lib* and *Dream Machines.* Redmond, Wash: Microsoft Press.

Raskin, J. (1989). The hype in hypertext: A critique. In *Proceedings of the Hypertext '87 Conference* (Chapel Hill, North Carolina, Nov. 13-15), 325-330.