The Web as an Object of Study

Steven M. Schneider
SUNY Institute of Technology
steve@sunyit.edu

Kirsten A. Foot
University of Washington
kfoot@u.washington.edu

August 25, 2003

Abstract

The ongoing evolution of the Web poses challenges for scholars as they seek to develop methodological approaches permitting robust examination of Web phenomena. Some of these challenges stem from the nature of the Web, which is a unique mixture of the ephemeral and the permanent. Increasingly complex Web applications that are altering traditional relationships between media form and content challenge the adaptation of traditional approaches to social research on the Web. Web-based media require new methods of analyzing form and content, along with processes and patterns of production, distribution, usage and interpretation. We identify three sets of approaches that have been employed in Web-related research over the last decade. Distinguishing between these approaches helps establish the trajectory of Web studies; highlighting the strengths and weaknesses of each focuses attention on the methodological challenges associated with the field of Web studies.

Key Words

Web studies, Internet research, methodology, Web archiving

Introduction

Over the last decade as email, the World Wide Web and various digital technologies have emerged scholars of new media have employed a variety of methodological strategies to explore social, political and cultural phenomena associated with the growth of these applications. Several recently published edited volumes highlight the range of methods employed in research regarding social dimensions of Internet technologies (Gauntlett, 2000, Jones, 1999, Mann, 2000, Howard and Jones, In Press). These collections, along with recent issues of scholarly journals, demonstrate that traditional methods of social research, such as ethnography (e.g. Markham, 1998, Hakken, 1999, Hine, 2000), textual analysis (e.g. Mitra and Cohen, 1999, Mitra, 1999, Crowston and Williams, 2000), focus

groups (e.g. Price and Capella, 2001, Stromer-Galley and Foot, 2002), surveys (e.g. Parks and Floyd, 1996, Smith, 1997, Schmidt, 1997, Yun and Trumbo, 2000) and experiments (e.g. Iyengar, 2002) have been adapted for use online in order to investigate both online and offline phenomena. In addition, some scholars have found it useful to employ Internet applications as bases for studies of purely offline phenomena (e.g. Witte et al., 2000). However, our focus is on the development of methods for studying social dimensions of the Internet itself, and in particular the Web.

As the Web has emerged as a distinct media form in the past ten years, it has increasingly been viewed as an object of study by social researchers. The ongoing evolution of the Web poses challenges for scholars as they seek to develop methodological approaches permitting robust examination of Web phenomena. Some of these challenges stem from the nature of the Web, which is a unique mixture of the ephemeral and the permanent. There are two aspects to the ephemerality of Web content. First, Web content is ephemeral in its transience, as it can be expected to last for only a relatively brief time. From the perspective of the user or visitor (or researcher), specialized tools and techniques are required to ensure that content can be viewed again at a later time. Second, Web content is ephemeral in its construction -- like television, radio, theater and other "performance media," (Hecht et al., 1993, Stowkowski, 2002). Web content, once presented, needs to be reconstructed or re-presented in order for others to experience it. Although Web pages are routinely reconstructed by computers without human intervention (when a request is forwarded to a Web server), it nevertheless requires some action by the producer (or the producer's server) in order for the content to be viewed again. In other words, the experience of the Web, as well as the bits used to produce the content, must be intentionally preserved in order for it to be reproduced (Arms et al., 2001). Older media – including printed materials, film and sound recordings, for example -- can be archived in the form in which they are presented; no additional steps are needed to re-create the experience of the original.

At the same time, the Web has a sense of permanence that clearly distinguishes it from performance media. Unlike theater, or live television or radio, Web content must exist in a permanent form in order to be transmitted. The Web shares this characteristic with other forms of media such as film, print, and sound recordings. The permanence of the Web, however, is somewhat fleeting. Unlike any other permanent media, a Web site may regularly and procedurally destroy its predecessor each time it is updated by its producer. That is, absent specific arrangements to the contrary, each previous edition of a Web site may be erased as a new version is produced. By analogy, it would be as if each day's newspaper was printed on the same piece of paper, obliterating yesterday's news to produce today's.

The ephemerality of the Web requires that pro-active steps be taken in order to allow a recreation of Web experience for future analyses. The permanence of the Web makes this eminently possible. Although saving Web sites is not as easy as, say, saving editions of a magazine, archiving techniques are evolving in such a way to facilitate scholarly research of Web sites. In distinction to other ephemeral media, the Web can be preserved in nearly the same form as it was originally "performed", (Kahle, 1997, Lyman and Kahle, 1998, Lyman, 2002), and analyzed at a later time. Web archiving enables more

rigorous and verifiable research, as well as developmental analyses that are time sensitive (e.g. Foot et al., 2003).

Approaches in Web Studies

Some of the broad questions currently under investigation by Web scholars include:

- What forms of communicative actions are being inscribed on the Web and how do they change over time?

- How do the actions of Web producers enable and/or constrain the potential actions of Web users?

- What kinds of user experiences are potentiated on and between particular Web sites?

- How are relations between Web producers, as well as between producers and users, enacted and mediated via Web texts and links?

These kinds of research questions, along with the increasingly complex Web applications that are altering traditional relationships between media form and content, challenge traditional approaches to social research. Web-based media require new methods of analyzing form and content, along with processes and patterns of production, distribution, usage and interpretation.

We identify three sets of approaches that have been employed in Web-related research over the last decade. These approaches are not necessarily mutually exclusive, and some studies cited below employed more than one approach. Distinguishing between these approaches helps establish the trajectory of Web studies; highlighting the strengths and weaknesses of each focuses attention on the methodological challenges associated with the field of Web studies.

The first set of approaches we identify employ discursive or rhetorical analyses of Web sites. This set of approaches is more concerned with the content of a Web site than its structuring elements. Studies employing these approaches focus on the texts and images contained on Web pages, and/or on Web pages/sites as texts in a Foucauldian sense (e.g. Baym, 1999, Benoit and Benoit, 2000, Sillaman, 2000, Warnick, 1998). Studies using a discursive/rhetorical approach, especially those that take broad views of what constitutes text, contribute significantly to our understanding of communicative phenomena on the Web. However, we contend that the classic arguments regarding the inseparability of form and content in traditional media are especially applicable to the Web, and thus studies of Web "content" that overlook the structuring elements of a Web page or site are also limited. Another limitation within this set of approaches is the paucity of analytical tools for making sense of the links among Web pages and between Web sites. Some studies of hypertext intertextuality include analyses of cross-site linking, (e.g. Mitra, 1999, Warnick, 2001), but most content-focused studies of the Web tend to reflect and perpetuate what we believe is an inadequate construction of the Web as merely a

collection of texts. As Burbules and Callister (2000 p. 83) observe, "people usually see points or texts as primary, and the links between them as mere connectives." We agree with their claim that links are "associative relations that change, redefine and provide enhanced or restricted access to the information they comprise," and we support the argument offered by Mitra, (1999), Berners-Lee (2000), Odlyzko (2001) and others that, on the Web, connectivity matters as much as content.

We characterize the second set of approaches as structural/feature analyses. Studies in this genre tend to use individual Web sites as their unit of analysis, focusing on the structure of the site, such as the number of pages, hierarchical ordering of pages, or on the features found on the pages within the site, for instance, the presence of a search engine, privacy policy, or multiple navigation options (D'Alessio, 1997, D'Alessio, 2000, McMillan, 1999, Hansen, 2000, Benoit and Benoit, 2000). Although understanding the structural and feature aspects of a particular site are important, our primary concern with these approaches is that they do not afford systematic analysis of an individual site's situatedness in the larger web, that is, the external pages to which it links, and the external pages which link to it. Another type of structural analysis employs computer-assisted macro-level network analysis methods for mapping linking patterns (e.g. Jackson, 1997, Park, In Press, Park and Thelwall, 2003, Rogers and Marres, 2000, Rogers and Marres, 2002). Studies of this type enable understanding of network structures on the Web, but inferring the meaning or "substance" of those network structures can be difficult to infer from large-scale mapping studies.

More recently, a third set of approaches to Web analysis has emerged that takes hyperlink relationality into account in more nuanced ways. We refer to this set of approaches for analyzing multi-actor, cross-site action on the Web as sociocultural analyses of the Web (see several examples in Beaulieu and Park, 2003). Lindlof and Shatzer (1998) point in this direction in their article calling for new strategies of media ethnography in "virtual space." Hine (2000) presents a good example of sociocultural analysis of cross-site action on the Web. Similarly, Howard's (2002) conceptualization of network ethnography reflects methodological sensitivity to processes of Web production. By appropriating the term "sociocultural" to describe this set of approaches we seek to highlight the attention paid in this genre of Web studies to the hyperlinked context(s) and situatedness of Web sites-- and to the aims, strategies and identity-construction processes of Web site producers-- as sites are produced, maintained and/or mediated through links.

Web Sphere Analysis

Our own work has benefited from the methodological groundwork established by our colleagues in Web studies. We are developing a multi-method approach called "Web sphere analysis" that enables analysis of communicative actions and relations between Web producers and users developmentally over time. We conceptualize a Web sphere as not simply a collection of Web sites, but as a hyperlinked set of dynamically defined digital resources spanning multiple Web sites deemed relevant or related to a central theme or "object." The boundaries of a Web sphere are delimited by a shared object-orientation and a temporal framework. Web sphere analysis is an analytic strategy that includes relations between producers and users of Web materials, as potentiated and

mediated by the structural and feature elements of Web sites, hypertexts, and the links between them (Foot and Schneider, 2002, Foot et al., 2003).

The most crucial element in this definition of Web sphere is the dynamic nature of the sites to be included. This dynamism comes from two sources. First, the researchers involved in identifying the boundaries of the sphere are likely to continuously find new sites to be included within it. Second, the notion of defining a Web sphere is recursive, in that pages that are referenced by other included sites, as well as pages that reference included sites, are considered as part of the sphere under evaluation. Thus, as a Web sphere is analyzed over time, (ideally via an archive that enables retrospective analysis), its boundaries are dynamically shaped by both researchers' identification strategies and changes in the sites themselves.

The Web sphere can function as a macro unit of analysis, by which historical and/or inter-sphere comparisons can be made. For example, the Web sphere of the 2000 elections in the United States can be comparatively analyzed with the electoral Web sphere of 2002 and those that develop in later years, as well as with electoral Web spheres in other countries. Alternatively and/or simultaneously, other, more micro units such as texts, features and/or links can be employed in analyses within a Web sphere (Schneider and Foot, 2002; Schneider and Foot, In Press).

Web sphere analysis is an analytic strategy that, fully implemented, includes analysis of the relations between producers and users of Web materials as potentiated and mediated by the structural and feature elements of Web sites, hypertexts and the links between them. In a nutshell, the multi-method approach of Web sphere analysis consists of the following elements. Web sites related to the object or theme of the sphere are identified, captured in their hyperlinked context, and archived with some periodicity for contemporaneous and retrospective analyses. The archived sites are annotated with human and/or computer-generated "notes" of various kinds, which creates a set of metadata. These metatdata correspond to the unit(s) and level(s) of analysis anticipated by the researcher(s). Sorting and retrieval of the integrated metadata and URL files is accomplished through several computer-assisted techniques. Interviews of various kinds are conducted with producers and users of the Web sites in the identified sphere, to be triangulated with Web media data in the interpretation of the sphere.

Conclusions

The emergence of the Internet, and especially the Web, has challenged scholars conducting research to both adapt familiar methods and develop innovative approaches that account for the unique aspects of the Web. The uniqueness of the Web includes both the nature of the communicative processes it engenders, and the challenges posed to create research repositories that allow robust analyses (that are representative and reproducible) to proceed.

Methodological innovations have emerged in correspondence with the properties of these new media applications. This analysis has highlighted some methodological trends. Earlier studies of the Internet tended to focus either on users and/or usage patterns, or on

media/production characteristics. Within the user studies genre, the predominant methods were various forms of textual/discourse analysis and participant observation, with online surveys and experiments emerging later. Overarching these trends is a shift toward methods that recognize the coproductive nature of new media-- and thus the duality of users and producers and the potential for digital media productions to be simultaneously inscriptions of communicative action and structures for action, especially on the Web.

The emergence of Web archiving techniques designed to facilitate scholarly analysis integrates researchers into archiving activities. Traditionally, the work of archivists proceeded largely independently of the scholars who would, at a later time, be expected to use the archived materials as the basis of research work. Given the cost and complexity of Web archiving, an alternative approach is emerging that attempts to integrate researchers into archiving activities. If scholars join with archivists to identify Web objects of interest, and to delineate strategies for building archives that support scholarly activities, the basis of future research efforts is likely to be enhanced and new methods for Web studies developed.

References

Arms, W., Adkins, R., Ammen, C. and Hayes, A. (2001), 'Collecting and Preserving the Web: The MINERVA Prototype', *RLG DigiNews*, April 15, 2001, 5(2): http://www.rlg.org/preserv/diginews/diginews5-2.html

Baym, N. (1999) *Tune in, log on: Soaps, fandom and on-line community,* Sage, Thousand Oaks, CA.

Beaulieu, A. and Park, H. W. (Eds.) (2003) Special Issue: 'Internet Networks: The form and the feel', *Journal of Computer-Mediated Communication*, 8 (4). http://www.ascusc.org/jcmc/vol8/issue4.

Benoit, W. J. and Benoit, P. J. (2000) 'The virtual campaign: Presidential primary websites in Campaign 2000', *American Communication Journal,* 3 (3). http://acjournal.org/holdings/vol3/Iss3/curtain.html#4

Berners-Lee, T. (2000) *Weaving the web: The original design and ultimate destiny o f the world wide web,* HarperCollins, New York.

Burbules, N. C. and Callister, T. A. (2000) *Watch IT: The risks and promises of information technologies for education,* Westview Press, Boulder, CO.

Crowston, K. and Williams, M. (2000) 'Reproduced and emergent genres of communication on the World Wide Web', *The Information Society,* 16 (3): 201-215.

D'Alessio, D. (1997) 'Use of the World Wide Web in the 1996 US election', *Electoral Studies,* 16 (4): 489-500.

D'Alessio, D. (2000) 'Adoption of the World Wide Web by American political candidates, 1996-1998', *Journal of Broadcasting and Electronic Media,* 44 (4): 556-568.

Foot, K. A. and Schneider, S. M. (2002) 'Online Action in Campaign 2000: An Exploratory Analysis of the U.S. Political Web Sphere', *Journal of Broadcasting & Electronic Media,* 46 (2): 222-244.

Foot, K. A., Schneider, S. M., Dougherty, M., Xenos, M. and Larsen, E. (2003) 'Analyzing linking practices: Candidate sites in the 2002 U.S. electoral Web sphere', *Journal of Computer-Mediated Communication,* 8 (4). http://www.ascusc.org/jcmc/vol8/issue4/foot.html

Gauntlett, D. (Ed.) (2000) *Web.studies: Rewiring media studies for the digital age,* Oxford University Press, New York.

Hakken, D. (1999) *Cyborgs@Cyberspace? An ethnographer looks into the future,* Routledge, London.

Hansen, G. (2000) 'Internet presidential campaigning: The influences of candidate internet sites on the 2000 elections', presented at National Communication Association, Seattle, WA, November.

Hecht, M. L., Corman, S. R. and Miller-Rassulo, M. (1993) 'An evaluation of the drug resistance project: A comparison of film versus live performance media', *Health Communication,* 5 (2): 75-88.

Hine, C. (2000) *Virtual ethnography,* Sage, Thousand Oaks, CA.

Howard, P. (2002) 'Network Ethnography and Hypermedia Organization: New Organizations, New Media, New Myths', *New Media & Society,* 4 (4): 550-574.

Howard, P. N. and Jones, S. (Eds.) (In Press) *Society Online: The Internet in Context,* Sage, Thousand Oaks.

Iyengar, S. (2002) 'Experimental Designs for Political Communication Research: From Shopping Malls to the Internet', presented at Workshop in Mass Media Economics, Department of Political Science, London School of Economics, June 25-26. http://pcl.stanford.edu/common/docs/research/iyengar/2002/expdes2002.pdf.

Jackson, M. (1997) 'Assessing the Structure of the Communication on the World Wide Web', *Journal of Computer-Mediated Communication,* 3 (1). http://www.ascusc.org/jcmc/vol3/issue1/jackson.html

Jones, S. (Ed.) (1999) *Doing Internet Research: Critical Issues and Methods for Examining the Net,* Sage, Thousand Oaks, CA.

Kahle, B. (1997) 'Preserving the Internet', *Scientific American,* 276 (3): 82-83.

Lindlof, T. R. and Shatzer, M. J. (1998) 'Media ethnography in virtual space: Strategies, limits, and possibilities', *Journal of Broadcasting and Electronic Media,* 42 (2): 170-189.

Lyman, P. (2002), 'Archiving the World Wide Web', *Building a National Strategy for Digital Preservation*, April, http://www.clir.org/pubs/reports/pub106/web.html

Lyman, P. and Kahle, B. (1998) 'Archiving Digital Cultural Artifacts: Organizing an Agenda for Action', *D-Lib Magazine*. http://www.dlib.org/dlib/july98/07lyman.html

Mann, C., & Stewad, F (2000) *Internet Communication and Qualitative Research Online: A Handbook for Researching Online,* Sage, London.

Markham, A. N. (1998) *Life Online: Researching Real Experience in Virtual Space,* Altamira Press, Walnut Creek, CA.

McMillan, S. J. (1999) 'Health communication and the Internet: Relationships between interactive characteristics of the medium and site creators, content, and purpose', *Health Communication,* 11 (4): 375-390.

Mitra, A. (1999) 'Characteristics of the WWW text: Tracing discursive strategies', *Journal of Computer-Mediated Communication,* 5 (1). http://www.ascusc.org/jcmc/vol5/issue1/mitra.html.

Mitra, A. and Cohen, E. (1999) 'Analyzing the web: Directions and challenges', in *Doing internet research: Critical issues and methods for examining the net* (Ed, Jones, S.) Sage, Thousand Oaks, CA, pp. 179-202.

Odlyzko, A. (2001) 'Content is Not King', *First Monday,* 6 (2). http://firstmonday.org/issues/issue6_2/odlyzko/index.html.

Park, H. W. (2003) 'What is Hyperlink Network Analysis?: New Method for the Study of Social Structure on the Web', *Connections*. 25(1), 49-62.

Park, H. W. and Thelwall, M. (2003) 'Hyperlink Analyses of the World Wide Web: A Review', *Journal of Computer-Mediated Communication,* 8 (4). http://www.ascusc.org/jcmc/vol8/issue4/park.html

Parks, M. and Floyd, K. (1996) 'Making Friends in Cyberspace', *Journal of Computer-Mediated Communication,* 1 (4). http://www.ascusc.org/jcmc/vol1/issue4/vol1no4.html.

Price, V. and Capella, J., N. (2001) 'Online Deliberation and its Influence: The Electronic Dialogue Project in Campaign 2000', presented at American Assocation for Public Opinion Research,, Canada, May.

Rogers, R. and Marres, N. (2000) 'Landscaping Climate Change: A Mapping Technique for Understanding Science & Technology Debates on the World Wide Web', *Public Understanding of Science,* 9 (2): 141-163.

Rogers, R. and Marres, N. (2002) 'French scandals on the Web, and on the streets: A small experiment in stretching the limits of reported reality', *Asian Journal of Social Science,* 30 (2): 339-353.

Schmidt, W. C. (1997) 'World-Wide Web Survey Research: Benefits, Potential Problems, and Solutions', *Behavior Research Methods, Instruments & Computers,* 29 (2): 274-279.

Schneider, S. M. and Foot, K. A. (2002) 'Online Structure for Political Action: Exploring Presidential Web Sites from the 2000 American Election', *Javnost (The Public),* 9 (2): 43-60.

Schneider, S. M. and Foot, K. A. (2003) 'Crisis Communication and New Media: The Web After September 11', In *Society Online: The Internet in Context* (Eds, Howard, P. N. and Jones, S.) Sage, Thousand Oaks, CA.

Sillaman, L. (2000) *The digital campaign trail: Candidate images on campaign websites.* Master's Thesis, Annenberg School for Communication, University of Pennsylvania, Philadelphia, PA.

Smith, C. (1997) 'Casting the Net: Surveying an Internet Population', *Journal of Computer-Mediated Communication,* 3 (1). http://www.ascusc.org/jcmc/vol3/issue1/smith.html

Stowkowski, P. A. (2002) 'Languages of place and discourses of power: Constructing new senses of place', *Journal of Leisure Research,* 34 (4): 368-382.

Stromer-Galley, J. and Foot, K. A. (2002) 'Citizen perceptions of online interactivity and implications for political campaign communication', *Journal of Computer-Mediated Communication,* 8 (1). http://www.ascusc.org/jcmc/vol8/issue1/stromerandfoot.html

Warnick, B. (1998) 'Appearance or reality? Political parody on the web in Campaign '96', *Critical Studies on Mass Communication,* 15 (3): 306-324.

Warnick, B. (2001) *Critical literacy in a digital era: Technology, rhetoric, and the public sphere,* Lawrence Erlbaum Associates, Mahwah, NJ.

Witte, J., Amoroso, L. and Howard, P. N. (2000) 'Method and Representation in Internet-Based Survey Tools: Mobility, Community, and Cultural Identity in Survey 2000', *Social Science Computer Review,* 18 (2): 179-195.

Yun, G. W. and Trumbo, C. (2000) 'Comparative Response to a Survey Executed by
     Post, E-mail and Web Form', *Journal of Computer-Mediated Communication,* 6
     (1). http://www.ascusc.org/jcmc/vol6/issue1/yun.html