Edited by
Jason Hughes

SAGE Internet Research Methods

Volume 4
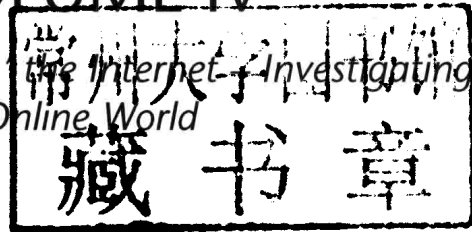
SAGE INTERNET RESEARCH METHODS

VOLUME IV

Research 'On' and 'In' the Internet: Investigating the Online World

Edited by

Jason Hughes

ⓈSAGE

Los Angeles | London | New Delhi
Singapore | Washington DC

SAGE

Los Angeles | London | New Delhi
Singapore | Washington DC

© Introduction and editorial arrangement by
Jason Hughes, 2012

First published 2012

SAGE INTERNET RESEARCH METHODS

Contents

Volume IV: Research 'On' and 'In' the Internet –
Investigating the Online World

The World of Web 2.0: Blogs, Wikis and Websites

Ted Gaiser and Anthony Schreiner

In the following pages we'll discuss the basics of web sites, weblogs (blogs) and wikis, and the ways in which they are being used for research purposes. In previous chapters we discussed technologies that allow users to send and receive information to and from a specific list of addressees. Web technology is different from email and messaging in that a document is published on the web and viewers themselves make the decision whether to access it. In the context of this book, the web is interesting in that it is both a medium for distributing information and a seemingly infinite source of data to research.

In its earlier days, that is, before roughly the year 2000, the web was mostly a publishing tool. Web servers were mostly in-house or at a commercial web service provider. A certain level of expertise was required to operate the web server and also to generate web content. Recent combinations of increasing network speeds, huge increases in disk storage capacity and software to ease the creation of web sites and web content, have made possible sites such as **MySpace, Facebook, YouTube, flickr** and ones like them, as well as blog and wiki sites. This has made it much easier for the average person (with access to a computer and the Internet) to express him or herself to a world-wide audience without having to go through a mediator such as a magazine or book publisher, or a television or radio station. This has empowered people, and is changing the relationship between commercial providers and consumers as all such people can now be providers as well as consumers (Wikinomics: Tapscott and Williams, 2006). A code word for this

new use of the World Wide Web is Web 2.0, which describes "this trend to enhance creativity, information sharing, and, most notably, collaboration among users" (http://en.wikipedia.org/wiki/Web_2.0, accessed on 6 May 2008). Of course, not all consumer-generated content is accurate or worthwhile, but then neither is all commercial content. The interactive web is in the process of significantly changing the nature of several industries, among them news, music, television and politics. These are all interesting areas for social science inquiry.

Most people today have some familiarity with website development, so we'll be brief. There are many ways to go about it. From most basic to more advanced, one can create the web documents by directly writing the HTML (HyperText Markup Language) code with a text editor, and transferring the document files to the web server document area. One can use web site editors such as **Adobe Dreamweaver** or **Mozilla SeaMonkey** to create the web content in a graphical WYSIWYG fashion, meaning that what you see on the screen is what you get when you print, and then transfer the file to the web server. Or finally the web site can be set up with a Content Management System (CMS), which is software on the web server that manages a whole web site. Contributors to the web site log in to the system, and are given access to some or all of the web site documents and, depending on permissions, are allowed to create and or update documents on the server. Examples of Content Management Systems are **Drupal, MediaWiki** and **Microsoft SharePoint.** All the big social networking sites like MySpace and Facebook are made possible by some type of CMS. The essential pieces of a content management system are an authentication system, a visual web document editor, and a storage scheme, which hides the actual form and location of documents from the user, and which is usually a relational database management system (RDBMS). A relational database is of interest to a researcher when he or she wants to retrieve content from a web site that he or she has access to. RDBMSs will be described in more detail in Chapter 8.

Weblogs (Blogs)

A particular form of user-generated web site is called a weblog or, in short, a blog. Blogs are web sites where an author creates documents such as diaries or commentary at (hopefully) regular intervals. Usually the author allows viewers of the site to post comments about the document, or engage in some type of online discussion. Subjects range from politics (e.g. www.democratic-conversation.com or dailykos.com) to personal and life events (e.g. www.triplebranch.blogspot.com) to consumer gadgets (e.g. gizmodo.com). Blogs often include images as well as text and dynamically link to other blogs and sites on the Internet. By convention, articles or posts are placed in reverse chronological order, that is, the most recent is at the top. Also by convention,

articles once posted, are not changed, though errata may certainly be posted later.

As Torill Elvira Mortensen notes, "the literature on weblogs is somewhat elusive, because the subject (weblogs), the academic publishing about weblogs, and both the academic and popular discussions about weblogs take place, to an extremely high extent, on the Internet" (in Coiro et al., 2008: 451). The practice of blogging itself is a "process of self-definition and a way of inscribing the individual on the global Net" (ibid.).

The characteristic of blog sites is that it is very easy for the author to create new content. So their usage is growing phenomenally. There are currently tens of millions of blog sites world-wide and more than a million articles are posted each day. Posting on a blog is less daunting to many people than writing a paper; partly because there is no editor (or teacher) to critique it, so blogs are often a way, particularly for young people, to exercise their literacy (Penrod, 2007). If there is a comment section to the blog, and if visitors to the blog use it, then that is a way for the author to get feedback, but it feels more like peer criticism than authoritarian criticism.

Blogs provide a significant amount of personal data. This is one of the reasons researchers find them of interest. Blogs represent easily accessible qualitative data for a researcher. While not necessarily focused on a particular topic, it can provide formidable data. It is important to note, however, that blogging is beginning to be questioned *as a means of* surveillance (Staples, 2000). Users place their personal data online, making these widely available to friend and foe. Whether or not people will continue to blog truthfully, whether they will reveal as much as they currently do, or whether it will evolve into a fictitious genre remains to be seen. For researchers, it is important to be clear for what purpose a blog is being used and in what ways the data might be misleading or intentionally fictitious.

Blogs, for example, have become the bane of a journalist's world in that they have made any 15-year-old child into new competition, providing a soap box for any and all opinions (Mortensen in Coiro et al., 2008). Editors note that it is difficult to remain on top of key stories, because blogs have enabled "netizens" to scoop the biggest stories. In a similar fashion, many anticipate that wikis will become the collaborative workspace of tomorrow, making them a part of corporate life. As they become a routine work tool, they will evolve into many other uses.

As a research tool, blogs represent qualitative data. They lend themselves to narrative and content analysis techniques specifically. In most cases, cutting and pasting is the easiest way to capture data for storage and analysis.

The preceding blog is fairly typical. It demonstrates a substantive, authored editorial on the topic of identity politics. It is then followed by a discussion with shorter contributions that reflect a kind of public dialog or debate. This type of data is useful for political analysis or other types of social science research. Managing the content for analysis in a program such

Example Sample Blog

March 08, 2007
Identity Politics Gone Wild

First we had the strikingly pale Bill Clinton proclaimed as America's First Black President by Toni Morrison as an odd reward for pandering to identity politics. Now, the *New York Sun* reports that the First Woman President may well *have a Y chromosome*, if John Edwards wins the White House:

Toni Morrison famously dubbed President Clinton America's "first black president." With that barrier broken, the comments of a prominent feminist are provoking debate about who may lay a similar claim to the title of America's first woman president.

The candidate being touted as a torchbearer for women is not Senator Clinton, but one of her former colleagues, John Edwards. At a rally near the University of California, Berkeley campus this week, a veteran of the abortion-rights movement, Kate Michelman, asked and answered the question she gets most frequently about her decision to back the male former senator from North Carolina.

"Why John Edwards, given the historic nature of our extraordinary campaign for the presidency this year with Hillary Clinton and Barack Obama and all the others?" Ms Michelman asked as she warmed up the crowd for Mr Edwards. "I've gotten to know a lot of political leaders over the years that I've been an advocate for women's rights. I know the difference between those who advocate as a political position and those who understand the reality of women's lives."

Compared to Mrs Clinton, Mr Edwards is short an "X" chromosome, but listening to Ms Michelman, that is easy to forget. "As a lawyer, as a senator, as a husband, as a father of two daughters, he understands the reality of women's lives. He understands the centrality of women's lives and experience to the health and well-being of society as a whole. . . . He understands that on an extremely personal level," she said.

Quite frankly, it shouldn't matter which chromosomes or skin pigmentation a candidate has on an individual basis. In a general sense, we want to ensure that the political process remains open to everyone, and diversity among elected officials gives us an indication of whether we're successful at it. If diversity is a goal in and of itself, as the Toni Morrisons and Kate Michelmans have argued in the past, then granting honorary minority status on rich, white men seems to be counter-productive.

It points out the silliness of identity politics in a comic way, as the final stage of the process. Identity becomes so much more important than actual policy

that candidates have to assume ridiculous poses as the most female of all candidates in a race, even while the race includes actual females and the claimant is a male. In Bill Clinton's case, the appelation belies the fact that actual black men and women ran for the office before he did, and did not win the nomination – and that he's not really black.

The first black president will be the candidate of African descent who wins the most Electoral College votes. The first female president will be the candidate with two X chromosomes who takes the oath of office on January 20 of the given year. Perhaps the Democrats might dispense with the gender and ethnic politics and just focus on policy instead.

Sphere It ⬛ ⬜ *1 blog reaction*
Posted by Ed Morrissey at March 8, 2007 06:02 AM

Listed below are links to weblogs that reference *Identity Politics Gone Wild:*

» *Thank You, Captain Ed* from Growing Old Disgracefully

All I hear in the discussions of the candidates is whether "Hillary is feminine enough" or whether Obama "is black enough". I still don't know Obama's foreign policy stance. I still don't know what or if Hillar . . . [*Read More*]

Tracked on March 8, 2007 08:22 AM

>Comments

"As a lawyer, as a senator, as a husband, as a father of two daughters, he understands the reality of women's lives. He understands the centrality of women's lives and experience to the health and well-being of society as a whole. . . . He understands that on an extremely personal level," she said.

Coming so soon after Ann Coulter's remark, this is just tooooooo good. If only Michelman had added "the importance of good hair" to her list of things that Edwards "understands". . .

Posted by: docjim505 ⬚ at March 8, 2007 06:38 AM

Now, what if that wannabe transgender City Manager in Florida goes on to bigger and better things? Could he/she be the "first woman president" if elected?

Posted by: BarCodeKing ⬚ at March 8, 2007 07:13 AM

"Perhaps the Democrats might dispense with the gender and ethnic politics and just focus on policy instead."

(*Continued*)

If properly explained, the Democrat policy is wildly unpopular with the electorate, which is why they resort to this sort of sham, hocus pocus redirection, rather than discussing what they actually want to accomplish (socialized everything), once elected.

Posted by: NoDonkey ⚘ at March 8, 2007 07:18 AM

Of course, Ms Michelman would probably be aghast at a Condi presidential run – even though she's both black and a woman.

And Mr Edwards, being identified as the "woman" presidential candidate won't help you.

Posted by: rbj ⚘ at March 8, 2007 07:22 AM

http://www.captainsquartersblog.com/mt/archives/009354.php,

accessed on 8 March 2007

as HyperResearch simply requires cutting and pasting the text into a word processor document file and inserting a tilde (~) between sections to enable the researcher to code each submission independently. If you have access to the content management system's back-end database on the blog server, then you can extract the posting and comments from the database with database tools. This may be preferable for large quantities of text, making it easier to prepare for analysis.

One of the challenges of the blogosphere is the sheer quantity of blogs. There are volumes of blogs in a number of different languages. Most researchers, therefore, will want to take advantage of some kind of service for assisting with their search for quality data, as well as managing the volumes of blogs they want to use in their research. One such service is Technorati. By their own statistics, Technorati (www.technorati.com) tracks over 112.8 million blogs and over 250 million pieces of tagged social media. As they note on their information site, "bloggers frequently link to and comment on other blogs, creating the type of immediate connection one would have in a conversation." Their service tracks these links, indexing tens of thousands of updates every hour. According to Technorati data, there are over 175,000 new blogs every day. Bloggers update their blogs regularly to the tune of over 1.6 million posts per day, or over 18 updates a second.

In addition to management tools, there are a number of search engines. One of the better search services is IceRocket (www.icerocket.com). It has a variety of criteria by which to search and various means by which to list a blog. Blogarama (http://www.blogarama.com/) lists over 82,000 sites to date, organized by subject heading, which points to another type of service.

Some services are helpful for organizational purposes, enabling a user to track blogs and their usage. One such service is del.icio.us, enabling users to organize and manage websites, blogs, and so forth, in a bookmark fashion similar to the way in which people are used to managing websites in a web browser.

In addition to capturing, organizing and analyzing data in existing blogs, a researcher could choose to create a blog. In the interactive environment of a blog, a researcher could manage a type of focus group by raising questions and challenging participants by responding to submissions. (Again, note the exchange in the example). Much like other electronic focus groups, this form of research presents challenges for focus group moderators (Gaiser, 1997). For example, it would be difficult in a blog setting for the moderator to have much control over the discussion, making it difficult to keep people on a particular task or discussion thread. Some researchers may be concerned about researcher bias, with so much authored information being the beginning of a discussion thread. The blog could present a kind of double-edged sword. Without substantial initial enticement to engage, participants may provide little follow-up, limiting the benefit of the blog environment for data collection. On the other hand, too much researcher involvement up front raises the issues of research bias. Lori Kendall notes, however, that "participant observation allows researchers to gain a better understanding of participants' ranges of identity performances and the meaning those performances have for them," suggesting that concern about researcher bias through participation in the study may be unfounded in online research (1999: 71). Regardless of any moderator challenges, however, researchers should not be deterred from attempting actively to capture the rich data accessible in a blog environment.

As with other internet technologies, blogs may be hosted on a personal web server or one belonging to your institution. This requires a working and accessible web server and blog management software, such as WordPress (wordpress.com), TypePad (www.typepad.com) or Moveable Type (www.moveable type.org). Blogs may also be hosted on a blog service, possibly using one of these same software management systems. Some of the biggest free (meaning advertiser-supported) hosted sites are Blogger (www.blogger.com) and Blogsome (www.blogsome.com). The social networking sites MySpace and Facebook also provide blogging facilities.

In every case, the author interacts with their blog through his or her web browser, and the procedure is fairly similar. Creating an account and a blog site on blogger.com is described in great detail in Sauers (2006). The main points are visiting the blogger.com site to first create an account, which is a Google account (and allows access to other Google services). Then you choose a name for your blog and a template which controls the appearance of the blog site. Once the blog site is successfully created, you can create

posts. You may either compose in raw HTML, or create with a visual editor. Among the many other options that can be controlled, are whether other contributors can create posts on the blog, whether or not comments will be moderated, and how often posts get moved off the main page and into the archived section.

Wikis

Wikis are similar to blogs in that they are user-generated web sites, but differ in that the document itself is allowed to be modified by multiple authorized users of the wiki site; it is a way of collaboratively creating and editing a document. A wiki is typically not a simple document like a journal article, but is often a complete website with different pages linked together. Wikis are generally not aimed at the general public, but are more often for the internal storage display of accumulated knowledge, and are therefore usually not fancy looking, but there is no reason they cannot be visually attractive. The name "wiki" comes from the Hawaiian word for fast, and is named after the Wikiwiki Shuttle at the Honolulu Airport.

The grand-daddy wiki is, of course, Wikipedia (www.wikipedia.com), an enormous collection of user-contributed knowledge. Anyone may request a log-in to Wikipedia and create an entry; there are currently over a million registered users. It is this openness that has made Wikipedia such a wide resource. There are some controls to limit vandalism of the site, but mostly the fact that every entry can be reviewed and edited by others generally pushes all entries towards greater accuracy. Comparisons between Wikipedia and standard encyclopedias show similar levels of errors (Giles, 2005). The advantage of Wikipedia is that, even if the errors are more frequent or greater, they will also be fixed more quickly in all likelihood.

The server technology behind wikis is the same as for blog sites. They require a log-in authorization system, graphically oriented web page editor, and a content management system, including a database back-end for information storage. Some wiki management products are MediaWiki (www.mediawiki.org/wiki/MediaWiki) (used by Wikipedia), and MoinMoin (moinmoin.wikiwikiweb.de). Wikis do not generally have fancy visual editors, but rely on a simplified mark-up language to create web pages that include a subset of the full capability of HTML mark-up.

Wikis are far less prevalent than blogs, perhaps because they seem more like work than play, and they have a more specific purpose, so there are fewer hosting sites for wikis. Among the ones that exist are Wikia (wikia.com) and WetPaint (wetpaint.com). It is also harder to search for wikis than blogs. One way is to use the Wiki Bus Tour, which can be started at http://en.wikipedia.org/wiki/Wikipedia:TourBusStop.
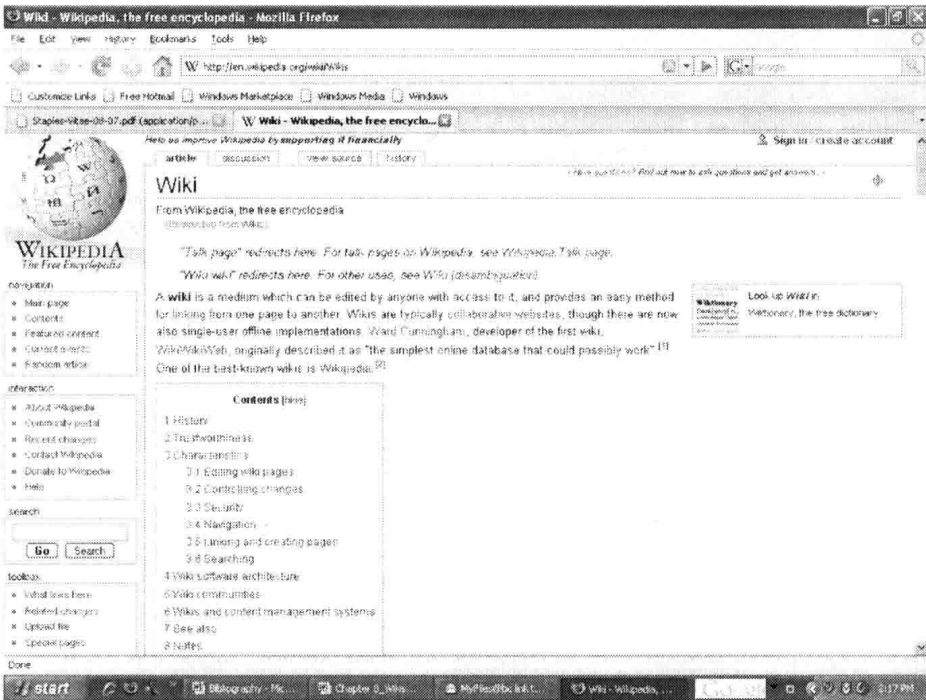
Feeds and Syndication

The very large number of blogs and dynamic web sites that are out there present a problem to researchers wishing to keep track of them. Certainly, web browsers have a bookmark feature that allows a browser to return quickly to a desired site, but it still requires the viewer to explicitly select the bookmark and go there. Increasingly, therefore, blog sites and other sites with frequently changing content provide what is called a "feed". More accurately such a web or blog page has a link on it which points to a special document, written in a form of Extensible Markup Language (XML) which can be imported by a feed reader or aggregator. These reader applications, with this information, know to check the web site automatically for new content and, if there is something new, present it to the viewer. There are a few different languages that provide the feed information: RSS which stands for one or all of Rich Site Summary, Really Simple Syndication or RDF Site Summary, and comes in several not entirely compatible versions, and also a separate language called Atom. But they are all dialects of XML and, for most people, what is in the feed link document does not actually matter; most feed readers will know what to do with any of them.

To provide a feed, the blog page or web page will usually display an icon that the user can click, either to get information about the feed, or actually to subscribe to the feed. These icons may vary, but are usually orange, and

may contain "RSS", "Atom", "RDF" or "XML". Additionally, modern browsers, including the latest version of each of Firefox, Internet Explorer, Safari and Opera, will detect the presence of a feed link, and put an orange icon like into the address field. Clicking on this icon will have different effects on different browsers. Firefox opens a window displaying the site in feed form, and gives the user an option of methods to subscribe. The "Live Bookmarks" option creates a feed icon in the tool bar, and subsequent articles from this site will appear there. It also gives the option of subscribing through a different application, which may be an email and news reader like Thunderbird, Office Outlook or Windows Mail. In a mail or news tool, the latest feeds will appear just as an unread email or news item, and can be marked or categorized in the various ways that mail tools are able. Lastly, Firefox gives the option of subscribing through Google or Yahoo, which may be useful if you use those sites as portals, and through the Bloglines (bloglines.com) site, which is a web site designed specifically for organizing feeds. Bloglines requires a free account, but has useful tools for organizing feeds. The other browsers – Internet Explorer, Opera and Safari – subscribe to feeds within the application itself. However, one last option is to find the feed icon in the document itself, not the one in the tool bar, and choose "copy link location", which is usually available in the mouse right-click option. The link location will be a URL pointing to an XML file. Then in the mail application, we will use Thunderbird as an example, select a new account of type "RSS News & Blogs". Then select "Manage Subscriptions", then "Add", and paste in the URL and select "OK". Your mail application will then be subscribed to this site's feeds.

Feeds are typically mostly text, but may in fact consist of audio and/or video, in which case they are known as Podcasts. The Apple iTunes application can be used to subscribe to audio and video feeds, and can save the feed to an iPod, hence the name. However, many applications that can present digital sound, like MP3 players, can play podcasts.

We end this chapter with a few notes on what a researcher needs to understand, should he or she wish to host a web site. Operating your own wiki or blog may be an attractive option for collecting data. However, running an interactive web site like a wiki or blog requires managing a web site with scripting enabled. This creates potential security vulnerabilities that your respective IT organization may resist. A decision will need to be made on who ensures the maximum possible security of the web server, including keeping up to date with patches, and also who repairs the damage in the event the server becomes compromised (see the section on Service Level Agreements in Chapter 2).

While relatively new to the world of research, wikis and blogs offer new opportunities for interacting with people on the web and gathering data. They need further exploration in regard to the value of the data generated in these environments. Currently, however, there are data and plenty of opportunity. It is our hope that these pages will inspire some to give them a try.

References

Coiro, J., Knobel, M., Lankshear, C. and Leu, D.J. (eds.) (2008) *Handbook of Research on New Literacies*. New York, NY: Lawrence Erlbaum Associates.

Gaiser, T.J. (1997) Conducting On-Line Focus Groups: A Methodological Discussion. *Social Science Computer Review*. 15 (2), pp. 135–144.

Giles, J. (2005) Internet encyclopedias go head to head. *Nature*. 438, pp. 900–901.

Kendall, L. (1999) The Conduct of Qualitative Interviews: Research Questions, Methodological Issues, and Researching Online, in S. Jones (ed.) *Doing Internet Research: Critical Issues and Methods for Examining the Net* (pp. 133–150). Thousand Oaks, CA: Sage Publications.

Penrod, D. (2007) *Using Blogs to Enhance Literacy: The Next Powerful Step in 21st-century Learning*. Lanham, MD: Rowman & Littlefield.

Sauers, M.P. (2006) *Blogging and RSS: A Librarian's Guide*. Medford, NJ: Information Today.

Staples, W. (2000) *Everyday Surveillance: Vigilance and Visibility in Postmodern Life*. Lanham, MD: Rowman & Littlefield.

Tapscott, D. and Williams, A.D. (2006) *Wikinomics: How Mass Collaboration Changes Everything*. New York, NY, Penguin.