

# Chapter One/First Draft

## What Is This World Wide Web?

by George Lorenzo

### INTRODUCTION

*This is the first draft of the first chapter of a book I am writing titled “SurfingThroughNoise™: Riding the Online Knowledge Wave.” The first chapter is an attempt to explain the most profound and important elements of the World Wide Web - perhaps an impossible task. It is meant to set the stage for the subsequent chapters of SurfingThroughNoise (STN), which takes for granted that its prospective readership is web savvy and wants to get a keener understanding of how the web is working and growing. In short, it is not an Internet guidebook.*

*STN is an overview of what the web is today and what it will possibly become. It is similar in its journalistic style and rhetoric to books that are based on solid research and timely information gathered through interviews with professionals in the field, along with references to and quotes from the pundits and mass media covering the Internet and the web. It describes a broad-ranging current state of affairs and its theoretical underpinnings and potential.*

*In addition, a value-add of STN is that its readers will learn how to surf through the web, and critically analyze it, in order to find information that will make them smarter human beings. It does not cover in any depth what's online in the world of pop culture, entertainment, shopping and travel. Instead, it covers topics that are of interest to anyone who wants to get a better sense of how the Internet and web are changing the way we acquire information and knowledge.*

## HOW BIG IS IT?

Nobody really knows the exact size of the World Wide Web. Much of the information about the size of the web and how fast it's growing is not consistent, and it does not look very authoritative. Near the end of 2006, Netcraft, an Internet services company based in Bath, England, claimed that the entire web had an annual growth of 30.9 million sites, achieving a new milestone of more than 100 million total websites. This statistic was heralded in an article by CNN.com, where Netcraft representative Rich Miller was quoted as saying that, out of the 100 million, only 47 or 48 million were actually active websites. In addition, two of the big reasons for the growth, as noted by Miller in the CNN article, were that it had become easier to create websites, as well as make money with them.<sup>1</sup>

Interesting! First, why is it that less than half of the sites are active? (More about that in another chapter.) Second, creating decent web pages is not an easy process. With all the relatively new features and functions one can add to a website today - such as discussion boards, blogs, wikis, RSS feeds, podcasts, multi media options, metadata, tagging, web services, and whatever you can discover about search engine optimization and marketing - it has become more difficult and complex than ever to create the modern-day equivalent of a moderately sophisticated website. You should also have some web design skills; be prepared to confront the multitude of choices related to e-commerce plans, web-building software and services, and hosting plans; and have a clear understanding of privacy and security issues? Plus, it is nothing less than an extraordinarily hard-working challenge to make even a small profit from any web-building and publishing efforts.

Yes, you (but definitely not everyone) can build a blog or a basic static website these days with free web-based templates and/or somewhat intuitive software (if you are web savvy). But if you're not a well-known and established brand, have strong marketing skills, are selling something at a real bargain that people want, have created that one-in-a-million better mousetrap, or are extremely lucky, forget about attracting more than your immediate family and close friends (on occasion) to your blog or website.

Call me a web skeptic. Although I love the web, I question it all the time. I question everything all the time. It drives my family nuts.

## CURIOSITY

For instance, it was easy to be skeptical about Netcraft. What is the company in Bath all about - a small town with a population of about 85,000 people located in the south of England, noted for its hot springs and Ro-

man Baths - and who runs it?

Of course, the first thing I did was visit the Netcraft.com website and click on the “About Netcraft” link. I was informed that this company is funded through “retained profit” and derives its revenue by providing network security services, research data and analysis on many aspects of the Internet and by accepting banner advertising on its websites. (I thought the “retained profit” wording was odd.)

It was also noted on the Netcraft website that it had clients in the UK, U.S., Europe, Middle East, Asia Pacific and Latin America, and many of them were big-name companies such as American Express, Deloitte & Touche, Paypal, Macromedia, Microsoft and others. In 1999, Deloitte & Touche ranked Netcraft as one of the fastest growing companies in the UK. Plus, Bob Metcalfe, who was noted on the Netcraft site as the “inventor of Ethernet,” had once called the Netcraft website “cool;” and Tim O’Reilly, founder and CEO of O’Reilly media, called the Netcraft website “the best known example of a site devoted to tracking technology on the Internet.” (I thought the “cool” notation was odd as well.) Conspicuously missing from all this self-promotional content was information about who ran the company. I’m not saying that Netcraft is not all the things that it says about itself. I’m only saying that if they would have told me who they are, I would have been more inclined to believe them.

The next step was to go to Network Solutions’ WHOIS online database, where anyone can discover who registered just about any domain name. Turns out Netcraft.com’s administrative contact was Mike Prettejohn, and it listed his e-mail address. Plus, through a Google search, an old press release dating back to February 2001 quoted Prettejohn as Netcraft’s president. That was good enough for me to shoot off an e-mail to him in which I introduced myself and explained that I am writing this book and would like to interview him by telephone in order get a clearer idea about his company and how he comes up with these figures.

As an education journalist I do this kind of inquiry all the time. I simply send out an e-mail query to a prospective interviewee with the subject line of “would like to interview you.” Roughly speaking, about 85 percent of the time I get an affirmative response and a telephone interview is ultimately conducted in a timely fashion. These interviews are basically the lifeblood of the work I do, and 99 percent of the time they are friendly, informative and thoroughly enjoyable. I’ve made numerous contacts (more than 700 over the past five years) and friends through telephone interviews that started with a simple e-mail query.

So, I was disappointed when I received a very polite thanks-but- no-

thanks e-mail response from Mr. Prettejohn, declining an interview.

I have yet to find any solid evidence about the total number of websites on the Internet. There are, however, lots of interesting and divergent statistics from numerous companies that study what websites get the most hits and visitors, what keywords are put into search engines, and much more. These constantly moving statistics are explained in more depth in subsequent chapters of this book. In the meantime, a solid method, borrowed from the schools of journalism, for deciding whether any website is authoritative and trustworthy is provided below.

The question of authority asks if the people providing information online are recognized experts. In other words, can their views can be considered valid, at least from the standpoint of their station in life, i.e. are they professionals in their respective field?

Trustworthiness takes the question of authority to the next level. Someone can be a recognized expert, but that does not mean that he or she is the kind of person who takes responsibility for his or her actions. Trustworthiness asks whether the people providing information online are credible. Do they possess the merits of believability and hence garner our trust?

## **APPLYING THE FIVE WS TO THE ONLINE WORLD**

Netcraft did not pass a simple, pragmatic website analysis, which is a twist on the journalism maxim commonly known as the Five Ws - Who, What, Where, When, Why, and How. (The Five Ws as a term and practice includes an unnamed H.) Here's how to apply the Five Ws to any website:

**Who:** Name the people or person in charge and provide their bios, with photos and an overview, as well as links to, downloadable files and/or web pages of past working experiences and client samples. (Some would disagree about the photos, saying that people make rash judgements based on mug shots.) Make it easy for your visitor to know who you are and be able to contact you or a representative of your business. There's an oversupply of websites that don't clearly tell you who runs the business, or these sites have a anonymous "contact-us" form that does not clearly identify who the form is actually being delivered to. This is one of the biggest credibility failures of many websites. If you can't tell people who you are, your readers might think you are hiding from something. There's more to the "who" function of any website. Explaining who you are should also include your version, or someone else's version, of what gives you the authority to present information that is credible and trustworthy. Netcraft.com failed the vitally important "who" test.

**What:** Any website should describe the nature of its owner's business in easy-to-understand terms. The other part of "what" concerns quality of content and design. Is the content well written? Is the website graphically consistent and easy to navigate through? The Internet industry term for this is "usability." In September 2005, the man known as "the king of usability" by Internet Magazine and "the guru of web page usability" by the New York Times, Jakob Nielsen, explained the irony of the web, noting succinctly that, while the primary purpose of the web is to provide information, it is also loaded with "bad content and a lack of information people need - either because it is not provided at all or because it is written in a poor, impenetrable style."<sup>2</sup> All you need is one search engine response to a query to realize that this still holds true today.

**Where:** Something is definitely amiss if there are not working e-mail addresses, a real physical address of where the business is located, and working telephone numbers listed in a easy-to-find spot on a website. Are you located in Silicon Valley, Silicon Alley (the Manhattan version) or Bath, England? Are directions to your place of business provided? Are you hiding?

**When:** Today, when it comes to information about the Internet, World Wide Web, and communications and information technology, in general, timeliness is vital, as everything changes so quickly. At the same time, we web surfers often fall into the trap of depending too much on the very latest information out there, when, in fact, there's plenty of valid and important, but older, information available online about any given topic, dating back as far as the web will take you. The important thing is that anything posted on a website should have some kind of time-stamp on it, so the reader understands the currency of the information being displayed and can then discern its applicability or non applicability to the task at hand.

**Why:** What are the motives behind the website owner's content? Is it clearly spelled out that the purpose of the content is to sell you something? Is the cost clearly noted, or is that information buried someplace at the end of a shopping-cart function? Is the content geared toward only providing information to its visitors in the spirit of sharing, or is there some other, not-so-evident, ulterior motive.

**How:** How was the information presented on the website discovered or created, and is it consistent with other information from other reliable sources? For example, I could not find any other authority claiming that there were 100 million websites. CNN, however, found it worthwhile to mention this figure like it was an absolute truth, as did hundreds of blogs.

One would think that generic, statistical information, such as the total number of websites on the Internet, would have some corroborative numbers elsewhere on the web, but, in this case, I could not find anything.

### **FINDING THE HARD TO FIND**

There are numerous websites that fail to meet some of the 5 Ws but are still credible and packed with trustworthy research-based information. Many are in the academic realm, created by educators who typically post their curriculum vitas, along with links to their scholarly writings, on very unattractive and poorly designed web pages that do not have decent metadata.

Metadata, for those who are not familiar with this word, is really not as technical as it sounds. Simply put, it is the identifying words and tags that are put inside a website's background code. Metadata is utilized by search engines to index and ultimately reveal search results when users conduct queries online. Metadata is not discernable on the web pages we see on our web browsers unless we go to the source view.

At a November 2006 Chronicle of Higher Education technology conference, Adam Smith, group business product manager for Google Book Search and Google Scholar, had this to say about metadata: "We love metadata; we'll take all you can give us, but it is a mess. When you really dig into it a little bit, parsing it and making sense of it for the older material is a disaster. But we are doing our best." Danielle Tiedt, general manager of Windows Live Premium Search added that "metadata is in a bad state."<sup>3</sup>

So, without going into extraordinarily technical details, the basic message is that there's plenty of intelligent life on the web that is not so easy to discover through the major search engines (e.g., Google, Yahoo, MSN and Ask.com in that order of popularity) because of the lack of good metadata.

### **ACCESSING THE SMART WEB THROUGH PROPRIETARY DATABASES**

Another enormous block of authoritative and trustworthy information available online resides within the full text of scholarly journals and various other publications that are accessible only through a paid subscription.

There are two ways to obtain access to the full text of such paid-subscription information: purchase it yourself or become a member or employee of an institution or company that provides access to such information through its library system. Academic libraries, for instance, subscribe

to proprietary databases and provide access to numerous paid-subscription publications to its students, faculty and staff. These patrons are authenticated users assigned with usernames and passwords that allow them into the institution's virtual libraries.

So, if you don't have lots of discretionary funds to purchase paid subscriptions, or you are not an authenticated user of some proprietary library system, the accessibility to some of the most authoritative and trustworthy information online is not at your fingertips.

### **THE DRIVE TO BRING MORE TO YOUR FINGERTIPS**

Both Google and Microsoft are developing online services that allow users to more easily gain access to such authoritative and trustworthy content. Google Scholar, for instance, has a Library Links program where partnering academic libraries are able to make their licensed resources available, through link-resolver technology, to those authenticated patrons who prefer to conduct their research through the Google Scholar interface as opposed to going directly to their institutional virtual library system interface.

Similar to Google Scholar, but not nearly as developed or sophisticated, is Microsoft's Windows Live Academic, which, according to Microsoft representative Tiedt, is focused on answering online search questions better by analyzing how its users conduct online queries. At the aforementioned Chronicle conference, Tiedt explained that Microsoft was working diligently on developing methods and functions that would bring more authoritative and trusted search results to its users.<sup>4</sup>

Both of these projects were in their early developmental phase at the time of this writing in early 2007, and it is anyone's best guess as to when they might be considered out of Beta and useful to any degree of acceptability by serious researchers.

Overall, if you depend only on search results from Google, Yahoo, MSN, Ask, or pretty much any of the search engines out there to answer your queries, you are not getting the most intelligent results. Most search engines simply give you extremely long lists of results that, for the most part, are a source of confusion and not very focused on showing you the most authoritative and trustworthy online information available today at the top of its search results.

### **SEARCH ENGINE BOOM**

In a subsequent chapter, more details about search engine technologies are provided. There are many search engine companies in early stages of

development that are striving to be alternatives to the big four - Google, Yahoo, Microsoft and Ask. Many of these newcomers' "main business proposition is to be bought by Google, or for that matter by Yahoo or Microsoft," wrote Miguel Helft in the New York Times. Helft also noted that, according to the National Venture Capital Association, venture capitalists, since early 2004 up through the end of 2006, invested about \$350 million dollars in 79 start-ups "that had something to do with Internet search."<sup>5</sup>

Other young search engine companies seem to be on a solid pathway for creating a search alternative that will be their own for years to come. One such company is Kosmix, whose co-founder, Anand Rajaraman, explained to me that they are not in business to be bought by Google. Instead, they are in business to provide a more sophisticated alternative to searching than any of the other search engine companies.

It's interesting to note that Kosmix was founded by two computer science Ph.Ds from Stanford University. The other founder of Kosmix is Venky Harinarayan. Google's Larry Page and Sergey Brin are both computer science Ph.D. candidates on leave from Stanford, and Yahoo's co-founders Jerry Yang and David Filo are on leave of absence from Stanford's electrical engineering Ph.D. program.

Ragaraman and Harinarayan have a strong background in building database technologies, having built online comparison-shopping technologies as co-founders of a company called Junglee, which was eventually acquired by Amazon in 1998 for \$250 million. Take a look at their website (Kosmix.com) and you'll see that their search results are quite unique.

## **MASS DIGITIZATION**

Google and Microsoft are also in the mass digitization/eBooks business. Google has "Book Search" and Microsoft has "Windows Live Book Search" - both of which were also in Beta in January 2007. Others, who came into this field well before Google and Microsoft, include the Open Content Alliance, Project Gutenberg, the Million Books project, the University of Virginia Electronic Text Center, and the Internet Archive (not an exhaustive list).

Some of today's Web pundits say that we have entered a new era in which it is possible to digitize all of the world's books into a universal, online accessible library. "Might the long-heralded great library of all knowledge really be within our grasp?" asks Kevin Kelly from Wired Magazine in a New York Times Magazine article. His answer is yes, and he provides his proof of concept, in part, by explaining how the digitiza-

tion process is being accomplished today:

Stanford University is scanning its eight-million-book collection using a state of the art robot from the Swiss Company 4DigitalBooks. This machine, the size of a small SUV, automatically turns the pages of each book as it scans it, at the rate of 1,000 pages per hour. A human operator places a book in a flat carriage, and then pneumatic robot fingers flip the pages - delicately enough to handle rare volumes - under the scanning eyes of digital cameras.<sup>6</sup>

There's controversy surrounding the mass digitization world, with questions about copyright infringement and who gets control of any such universal library. The Associated Press reported that there's a philosophical debate concerning Google, a commercial entity that is scanning 3,000 books per day, and possibly controlling mankind's accumulative knowledge. The article quoted Brewster Kahle, founder of the Internet Archive, as saying that "they [Google] don't want the books to appear in anyone else's search engine but their own, which is a little peculiar for a company that says its mission is to make information universally accessible."

In a subsequent chapter, mass digitization and eBooks are explored in greater depth.

## **THE PARTICIPATORY WEB**

Then, of course, there's the "participatory web," also referred to as "user-generated content," "we media," "social media," the "democratized web" and a variety of other names. Most of the participatory web also lacks good metadata and hence is also difficult for search engines to discover. This can be considered a good thing, however, because most of the content on the participatory web is pure junk.

Nonetheless, the participatory web had a banner year in 2006. The Time Magazine December 25, 2006 issue cover story was devoted to the participatory web, naming "You" the Person of the Year, and calling "You" the Web 2.0 revolution, where "the stupidity of crowds as well as its wisdom" are harnessed online. The TV show *20/20* was on Time's heels with a special two-hour broadcast on December 29, titled "Caught," which featured "the craziest, funniest, most dramatic and most compelling images captured this year and shared online."

## **THE BIRTH OF NEW VOICES**

Just what is the participatory web? According to Carleton College Cin-

ema and Media Studies Department Professor John Schoot, who teaches an innovative six-credit course titled “Participatory Media,” it’s where anybody can gather, produce and publish their knowledge about anything to the world through a wide variety of new media, such as weblogs, photo blogs, podcasts, and video blogs. It’s the ability to find, collect, archive, share and remix audio, video and images online in a new Do-It-Yourself (DIY) culture.

What are the new realities of the participatory web? There are two schools of thought. One is that the participatory web is like the Tower of Babel and only adds to an already overabundance of irrelevant, hard-to-comprehend information published online. The other is that the participatory web has become the home for new individual voices and like-minded communities of interest that are catalyzing meaningful cultural and political change, with the same, or greater, level of credibility and importance as professional mass media.

Some of the literature about these two realities have strong voices. For instance, Jaron Lanier, computer scientist and Discover Magazine columnist, referred to the participatory web, ala wikis and other forms of social networking, as a new kind of social collectivism driven by a hive mind that is dangerous, stupid, boring, and, at times, capable of lowering the overall expectations we hold for individual human intellects.<sup>7</sup>

Best-selling author Steven Johnson added his take on Lanier’s point of view, when he wrote that

A swarm of connected human beings is a fantastic resource for tracking down software bugs or discovering obscure gems on the web. But if you want to come up with a good idea, or a sophisticated argument, or a work of art, you are still better off going solo.<sup>8</sup>

Yochai Benkler, Yale law professor, wrote a 515-page book about the participatory web (and much more) titled “The Wealth of Networks: How Social Production Transforms Markets and Freedom.” In a nutshell, Benkler asserts that we are in the midst of a new information age that has given us the freedom to actively participate in a networked information economy, i.e. the participatory web, that is not motivated by financial profit or managed by an industrial complex.

This new freedom holds great practical promise: as a dimension of individual freedom; as a platform for better democratic participation; as a medium to foster a more critical and self-reflective culture; and, in an increasingly information-dependent global economy, as a mechanism to achieve improvements in human development everywhere.<sup>9</sup>

The participatory web is also explained and discussed to a far greater extent throughout this book.

## **RECAP**

So far I have painted this small picture of the web as being hard to measure in size, loaded with hard-to-find authoritative and trustworthy content, packed with both stupidity and wisdom, and something that continues to grow at an enormously fast rate through mass digitization and through the adoption of new media. I've also talked briefly about search engines and have provided a methodology, that is not rocket science, for discerning what is valid information online. All of this is only scratching the surface of the web.

Some very important elements of the Internet, the web and today's information age that have not yet been mentioned, but are covered throughout STN, include such terms that you may or may not be familiar with: mashups, mobile computing, social networking, cyberinfrastructure, web services, virtual worlds, grid computing, social networking and bookmarking, content aggregators, podcasting, RSS feeds and Ajax and Atom, bit torrent, Library 2.0, the Long Tail, collaborative authorship, and citizen journalism. Plus, there are many other terms and topics of interest related to the information explosion spreading online that I have yet to discover or explore. Each day I am surprised by some new development or turn of events that looks to have the potential of bringing about dramatic change.

## **ON NOISE**

A negative side effect of this researching, interviewing, learning and reporting experience has been that I am frequently overwhelmed, as I would imagine anyone trying to harness and better understand the web would find themselves. As I continue to surf through the web, I find myself, at times, holding up my head with my hands covering my ears, like I'm attempting to cover up some "noise." In short, the web, with all its new implications that change with the click of a mouse overnight, has become a morass of incomprehensible noise, a cacophony of websites and web services. My goal now is to somehow make it quieter, with a tonal quality that I can control and listen to in comfort, similar to turning the volume down a few notches on the stereo or radio, or, better yet, the MP3 player.

Noise, as defined through a "define: noise" prompt in Google brought me this response, among many: "Noise is incomprehensibility resulting

from irrelevant information or meaningless facts or remarks.” Additionally, the 2e definition of noise from Merriam-Webster’s Collegiate Dictionary, 11th Edition, is “irrelevant or meaningless data or output occurring along with desired information.”

Learning how to find and analyze the desired information online - the kind of information that can help solve problems and challenges, answer our deepest questions, and perhaps bring about some positive change in our culture and politics - is what every web-savvy citizen needs to pursue more ardently than ever before. The web can provide us with what we need to know, and, from an historical perspective, what we have never been privy to see before. The ocean we call the web continues to expand ferociously into something we cannot accurately predict. But two elements of the web are certain: there’s a strong cross current of garbage and misinformation, and a strong cross current of wonderful gifts of knowledge and accurate, useful information at our fingertips. At the risk of sounding corny - “Surf’s Up, Dude,” - let’s ride the online knowledge wave, stay balanced, learn how to avoid nasty undertows, know where we are at all times and reach the shoreline safely so we can hop on the next wave.

## CHAPTER ONE ENDNOTES

1. Marsha Watson, "Web Reaches New Milestone: 100 Million Sites," CNN.com, November 1, 2006, <http://www.cnn.com/2006/TECH/internet/11/01/100millionwebsites/index.html>
2. Business Week Online, Best of the Web "Online Extra: Jakob Nielsen on the Unwieldy Web," Business Week, September 26, 2005, [http://www.businessweek.com/magazine/content/05\\_39/b3952418.htm](http://www.businessweek.com/magazine/content/05_39/b3952418.htm)
3. The Chronicle of Higher Education Technology Forum, "Strategies for Campus Leadership," Lake Las Vegas, November 13, 2006.
4. Ibid.
5. Miguel Helft, "In Silicon Valley, the Race Is On to Trump Google," The New York Times, January 1, 2007.
6. Kevin Kelly, "Scan This Book!," The New York Times Magazine, May 14, 2006.
7. Jaron Lanier, "Digital Maoism: The Hazards of the New Online Collectivism," Edge, May, 30, 2006, [http://www.edge.org/3rd\\_culture/lanier06/lanier06\\_index.html](http://www.edge.org/3rd_culture/lanier06/lanier06_index.html)
8. Steven Johnson, "Digital Maoism," The New York Times Magazine, 6th Annual Year in Ideas, December 10, 2006.
9. Yochai Benkler, *The Wealth of Networks: How Social Production Transforms Markets and Freedom* (New Haven, London: Yale University Press, 2006), 2.

*George Lorenzo is writer, editor and publisher of Educational Pathways and president and CEO of Lorenzo Associates, Inc. For more information about SurfingThroughNoise, please visit <http://www.edpath.com/stn.htm>*

*To make comments or suggestions about SurfingThroughNoise, please visit the STN blog at <http://georgelorenzo.blogspot.com/>*

**Copyright. © 2007 by George Lorenzo  
All Rights Reserved.**