

BY RALPH WESTFALL

If Your Pearls of Wisdom Fall in a Forest...

THE IDEA OF DOING THINGS THAT CAN IMPROVE something is an extremely popular concept in American culture. For example, I found the phrase, *make a difference*, in 129 million Web pages in a Yahoo! search in July 2009. The concept typically applies to impacts at local levels because, until recently, few people had opportunities to do things that could positively affect substantial numbers of people throughout the country in which they live or even throughout the world.

After dot com bubble burst, many investors found that the slogan that “the Internet changes everything” did not apply to many of the requirements for having a successful business. However the Internet really does provide opportunities for those who create knowledge to share it with more people who can use it to advantage, and share it more quickly than through other means.

Of course, this potential capability to share materials is limited by people’s abilities to find useful information in the billions of pages on the Internet. Yahoo reported indexing over 19 billion pages in 2005. (Perhaps responding to criticism of the estimates,¹ the leading search engine companies no longer publish counts of pages indexed. Although no one really knows how big the

“haystack” is, the simile about “finding a needle” is quite applicable to the Internet.)

Nevertheless there are two aspects of innovative materials that work in favor of their being found:

- ▶ The goals of search engine companies.
- ▶ The nature of knowledge.

Search Engine Company Goals and Characteristics

The very reason for the existence of search engines is to provide materials that are of value to those who are looking for things they need to find. The companies spend astronomical amounts on research to make their technologies accomplish this even better. Google spent over \$2.8 billion on R & D in 2008⁴ and Microsoft in particular is also allocating large amounts on research related to its Internet activities.⁶

What does this mean to a content provider? It means that if you have something on a Web page that is of value to other people, the search engines want to help them find it. Since the typical searcher pays the most attention to the top three items³ in a search listing, the search engine companies would like their algorithms to make your page show up in the top 10 or higher for people who are looking for something very much like that.

However search engines work on the basis of algorithms that relate the words being searched for to the content of Web pages. These algorithms are not capable of reading the minds of the searchers so they are dependent on the content of Web pages and other aspects including those of other pages that link to them. Therefore pages that adhere to patterns of writing that have historically facilitated the communication of what a page is really about generally show up higher in search listings than pages with comparable content that do not communicate this as well.

Search engine optimization (SEO). By now a lot of people realize that communication style can substantially affect a Web pages position (ranking) in search listings. This has led to the de-

velopment of a new employment field of SEO specialists and a lot of spam from people who say they can guarantee a top 10 listing. However the techniques that they provide are not particularly complicated or unknown. There are many Web sites that identify the main points for example, Search Engine Watch.⁸

The premise that SEO can make a page rank much higher than it really should goes against the search engine companies' goals of making their results as useful as possible. A significant part of their research involves improving their algorithms to weed out pages whose rankings have been manipulated to be higher than they deserve to be. Although Google has been occasionally exploited—such as, by Web pages that download malware² – it is quick to eliminate such pages from its rankings. Therefore such deceptive techniques are useless for those who want their material to be found over the long term. On the other hand, an approach designed to be consistent with these companies' goals of providing good results should be very viable over time.

Textbook metaphor. The use of metaphors in graphical user interfaces (GUI) has made computers more usable for a much larger population than those who were using command line interfaces. The metaphor of publishing in general and textbooks in particular can also be helpful in understanding how to make a Web page show up substantially higher in search engines than it would otherwise.

► The **title** of a textbook should communicate the essence of what it is really about, in a few words. This title goes into prominent places: the front and spine of the book, in one or more of the first few pages and possibly at the top of most pages. Translating this into HTML, the <title> tag, whose content shows in the bar at the top of the browser, should contain the “key words” that people would use to search for precisely what a page contains.

► Textbooks contain introductory material at the front and often have summary material at the start of chapters. Corresponding to this, material at the start (top) of a Web page that **summarizes** the content in more detail will enhance the communication of the whole page's content. (Shifting the metaphor,

an abstract at the start of an academic publication does the same.)

► A textbook has **headings** inside—to introduce the book and individual chapters. These are followed by a graduated pattern of sub-headings that get smaller as they introduce subtopics within a chapter. These headings correspond to the <hx> tags in HTML (<h1> largest and <h6> smallest). Using key words appropriately in headings will improve search engine rankings. It will also make it easier for people to find what they are looking for in your pages and increase the probability that they will be able to apply it to their needs.

► The “**density**” of key words is the percentage of a keyword to the total number of words in a page. If a page is really about something, words that are related to it will likely occur in it disproportionately often, just as they would in a textbook or in its individual chapters.

► In textbooks, **images** have captions that summarize their content. In HTML, the image tag has an alt= property that pops its text up when the viewer runs the cursor over the image in most browsers (the title= property does the same in Firefox). The contents of the alt= attribute also show instead if the image doesn't load. For visually challenged page visitors, the text of these alt= attributes can be converted to audible words. Search engines factor these words into their rankings. Therefore improving the communications capabilities of the page for a specific class of users can make it easier for others to find it. The same principle should also apply to tables (<table>) and lists (,) in HTML.

► Most Web pages have **links** to other pages. To be most effective, the links should contain words that clearly communicate what the linked pages are about. Linking to other pages that cover the same or related topics is a way of adding value for the viewer and also enhancing the page's search rankings.

As a cross validation of this approach, note that an empirical study¹¹ found that key words in the<title>and increasing key word density in the text separately improved search engine rankings, and even more so in combination.

Computer programming metaphor. An important concept in computer

programming is “self documenting code.” Programmers should use variable names and comments to make their code easier to understand. This metaphor can be easily adapted to search optimization. Behind the Web page is HTML code, and it can be made easier to understand. For example, most pages have images. They have file names that can be created to reflect the content of the images. The directory paths and file names of other pages in a Web site can also be chosen to help communicate the content of the page and site.

Page developers can also provide a description of the page contents and a list of key words in <meta tags. Although these tags have been manipulated so much that some search engines now ignore them, the description= text may show in the text adjacent to the link in search listings. If worded appropriately, this description could encourage searchers to enter the site.

Inductive Analyses. One of the big problems with typical SEO information is that it focuses on what seems to work rather than on the goals of the search companies. The companies are very dynamic in responding to attempts to manipulate their rankings, so things that worked in the past (such as, <meta > tags) may not continue to be beneficial in the future and might even become counter-productive. Good faith efforts to improve the communications capabilities of pages should be more successful over the long-term than following “hot tips” on this subject.

On the other hand, doing an inductive analysis (see sidebar) could be a good exercise for anyone interested in learning more about this topic. If used with a reasonable sample of Web pages comparable to an entry page that you want to optimize, it will probably corroborate the findings reported above. It might even suggest specialized techniques that might not be highly correlated with communications capabilities. However keep in mind that search engines factor in other issues that are external to pages (especially incoming links), so that the findings from a small sample may reflect random variation rather than being true indications of the impact of any novel technique identified through the analysis.

Conducting an Inductive Analysis of Search Engine Optimization Factors

- ▶ Search using words that you think should be good for finding your pages.
- ▶ Click the links for the top ten or so pages in the listing for various search words and phrases.
 - If the page seems comparable in size and content, use View>[Page] Source in the browser to see its HTML code.
- ▶ Use Edit>Find to locate occurrences of the keywords that led you to the pages. Tally the frequency of each of these words, and note their positions within the page and the HTML tags surrounding them.
 - If you are not familiar with HTML, you could view the effect on the corresponding text in the browser. However you will still need to check a HTML reference to identify items that are not seen in the browser, including words in URLs that links refer to, alt= attributes associated with images, <meta tags, etc.
- ▶ After completing this process with 5 to 10 highly ranked pages, you should have a fairly good idea of what you could do to your page to enhance its visibility to search engines on those words.

of each publication and also links to the full text of ones that are available online. Further down there is a link to a prominent researcher in the category, and other links that relate to the category. Based on this pattern each category page is highly focused on its particular topic, which leads to a relatively high proportion of key words relevant to its subject. This improves both the page's search engine ranking and also the rankings of pages that it links to.

Knowledge and New Knowledge

Specialization is one of the ways that societies cope with large amounts of information. The more specialized knowledge is, the fewer specialists there are that are concerned with it. And the word "new" implies different or unique.

These characteristics work together with the goals of the search engine companies to make it easier to make useful materials more readily available to those who could benefit from them. A person that generates new knowledge, particularly if it is specialized, is not competing with billions of Web sites. The competition will be very limited. It is also likely that many of the pages that are relevant have not been enhanced to improve the way they communicate what they are about to search engines, which gives a further advantage to those who know how to do so.

Note that for academics, Web publishing provides additional benefits. Steve Lawrence of the NEC Research Institute found that, from a sample of "119,924 conference articles in computer science and related disciplines," articles available online were cited an average of 7 times, in contrast to only 2.74 times for ones that were not online.⁵

Web Page Case Study

My experience indicates that search engine optimization (SEO) is relatively easy if a person is trying to improve the rankings of a page on searches that look for that type of content. My 1997 conference paper⁹ about the generally poor quality of research concerning telecommuting-related productivity increases was published on a server of one of the conference sponsors. In 1998 the paper was showing relatively highly on searches for telecommuting AND productivity in Alta Vista. Therefore I studied the limited search engine optimization ma-

Creating Your Own Portal

What made Google a better search engine from the start was that it evaluates not just a page's content but also the number of links to it from other pages. It also evaluates the quality of the incoming links by how many pages link to them in turn. If a person creates a good Web page on a topic and the content is structured so as to receive good search engine visibility, others will find it and link to it, and the page will then rank even higher in searches. However authors can accelerate this process by creating pages that link to their other pages. Note that this has to be done in a natural way that makes sense, rather than being a fraudulent attempt to manipulate search engine rankings.

I created a Web site¹⁰ in 2001 to be a portal to all my research papers and related writing. The entry page (shown in

Figure 1) categorizes the publications into four categories. Each category link on the portal page goes to a page with publications on that topic. There are category pages for telecommuting, information systems education, IS research relevance and end-user computing.

The entry page links to authors with large numbers of publications in the whole information systems field. Each category page has links to other authors in its topic. (If more authors follow this pattern, cross links between them would be beneficial.)

The pages all follow a similar pattern that could be used as a template. Links across the tops and bottoms of the pages go to category pages. Each category page has a <title> that represents the topic that it covers. At the top of each category page is a list of publications within that topic, with brief synopses

Figure 1. Personal Research Portal



terials available at the time, and also did inductive analyses of the HTML content for some high-ranked pages with a view to further improving its position.

Based on these analyses, I tweaked the paper for search engines. For more than five years it has been in the first 10 results (after the sponsored listings) in Google and a number of other search engines. Searches on July 22, 2009 found the page ranked number one in Open Directory, Bing, and EntireWeb; 2nd in Ask/Teoma, All the Web, and Yahoo; and 3rd in Google for the individual words telecommuting productivity (without using quotes around them).

Note that these high ranking are partly because relatively few people are interested in *telecommuting productivity*. The optimizations would not have placed the paper so highly for a broader topic. On the other hand, the narrow focus correlates well with specialized information, including academic research findings, because this kind of content is usually concentrated in a relatively limited number of Web pages.

The high rankings have been maintained in part because the optimization techniques are designed to more clearly express the content of the page. There are no keywords in invisible text (font color the same as the background) or text in a very small font, or repetition of words in a fashion that is inconsistent with conventional writing styles. Such misleading practices are known as “search engine spamming.” If they are detected by search engine algorithms, a page’s ranking will go down rather than up. It also is likely that the ranking of this page has been supported by including it among my telecommuting publications when I created the research portal described above.

Epilogue: AACSB and “Practical Research”

The primary accrediting agency for business schools, the Association to Advance Collegiate Schools of Business International, recently proposed that business schools evaluate their faculty on the practical impact of their research on the organizational world in addition to traditional academic publishing.⁷ Since the majority of information systems programs (MIS, CIS, IS) are housed in colleges of business, this would be a good time for fac-

ulty in these programs to start making their research findings easier to find on the Internet to increase the probability that their research will have an identifiable practical impact. **C**

References

1. Cheney, M. and Perry M. A comparison of the size of the Yahoo! and Google indices. 2005; http://www.mostpopularsites.net/MPS_News/Search_Engine_Facts/A_Comparison_of_the_Size_of_the_Yahoo!_and_Google_Indices/.
2. Eckelberry, A. and Thomas, A. HEADS UP: More Google poisoning on the way? *Sunbelt Blog*, (Nov. 29, 2007); <http://sunbeltblog.blogspot.com/2007/11/heads-up-more-google-poisoning-on-way.html>.
3. Eyetools, Enquiro, and did-it uncover search’s Golden Triangle. *Eyetools, Inc*; http://www.eyetools.com/inpage/research_google_eyetracking_heatmap.htm.
4. Google, Inc. Form 10-K for the fiscal year ended Dec. 31, 2008. 2009; http://investor.google.com/documents/2008_google_annual_report.html.
5. Lawrence, S. Online or invisible. *Nature* 411, 6837 (2001), 521; <http://citeseer.ist.psu.edu/online-nature01/>.
6. Murphy, K. Microsoft sees Google’s \$500m, raises \$500m. *Computer Business Review*, (May 5, 2005); http://Web.archive.org/Web/20070817091416/http://www.cbronline.com/article_news.asp?guid=3D810B1B-BBE0-482D-A81C-DBE60BAB97C4.
7. Practically irrelevant? What is the point of research carried out in business schools? *Economist.com*, (Aug 28, 2007); http://www.economist.com/business/globalexecutive/displaystory.cfm?story_id=9707498.
8. Sullivan D. Intro to search engine optimization. 2007; <http://searchenginewatch.com/showPage.html?page=2167921>.
9. Westfall, R. D. Does telecommuting really increase productivity? Fifteen rival hypotheses. AIS Americas Conference (Indianapolis, IN, 1997), 405-407; <http://www.empirenet.com/westfalr/prdctvty.html>.
10. Westfall, R. D. Information Systems research. 2001; <http://www.csupomona.edu/~rdwestfall/ais/>.
11. Zhang, J. and Dimitroff, A. The impact of Web page content characteristics on Web page visibility in search engine results (Part I). *Information Processing and Management* 41, 3, (May 2005), 665–690. <http://linkinghub.elsevier.com/retrieve/pii/S0306457303001122>.

Ralph Westfall (rdwestfall@csupomona.edu) is a professor in the Computer Information Systems Department at California State Polytechnic University in Pomona, CA.

© 2009 ACM 0001-0782/09/1100 \$10.00

Copyright of Communications of the ACM is the property of Association for Computing Machinery and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.