

Seek and ye shall find

Searching Beyond Google

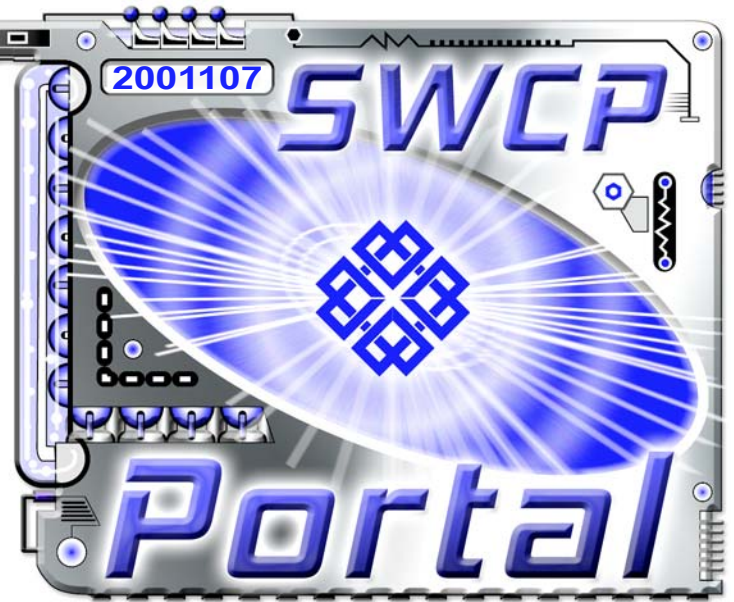
Now virtually the default search engine for the entire web, **Google** towers over all others. Scope, content, useful rankings and results, plus breadth of coverage and sheer speed, are among the many reasons for the growth and popularity of this huge Internet index.

But for all its virtues, Google is not the only search engine out there, and might not be the best one to use for your every inquiry. There are numbers of sophisticated and surprising alternatives, and you don't need Google to find them either.

The first reason to consider other online tools is **privacy**. Google's vast reach can be intimidating. It freely collects data from users as well as the World Wide web, and the results are easy to see. You may have noticed that many pages can be slow to load while they exchange data with Google or that ads they scatter across the web seem aimed directly at you.

If so, you're not imagining: Google's ads are highly personalized. The ones you see are chosen by your interests expressed not just by ads you click, but what you search for and the sites you visit. Such targeting is both convenient to users and it moves products. But there is an obvious downside. For instance, if there is a potentially sensitive condition that you looked up, it's possible that you might find yourself later running into off-putting references everywhere you surf.

Implications can be more subtle or far more severe than personal embarrassment, however. At the far extreme, scans of search histories are now often included in felony criminal investigations. Moreover Google, like **Facebook**, seems to regard individual privacy as somehow obsolete. As Eric Schmidt, Google's CEO, said in one interview, "If you have something that you don't want anyone to know, maybe you shouldn't be doing it in the first place."



How Search Engines Work

Early indexes, such as **Yahoo**, www.yahoo.com, in its infancy, relied entirely on people power. The directories were produced by humans laboriously clicking links and making lists. But the web's information explosion soon made it impossible to keep up.

So developers focused on automatic ways of scouring the Net for information. By using "spiders" and "bots"; programs that scan webpages for links, keywords and unseen data, they could generate and constantly update huge tables of data about which data was located where. Massive, ever-growing server farms are required to handle it all and stay apace.

Google cleverly ranks pages not so much by what they contained but by the number of links to them. The more links to a page, the more popular it must be and the higher it moves on the list. Of course, the more popular a page becomes and the higher its rank, the more links it tends to get. This feedback loop means that if a page is not on Google, it's like it doesn't exist. Showing up at the end is almost as bad.

Many webmasters deliberately design their sites to be as attractive to the web-crawlers as to human readers, if not more. The practice, called "**search engine optimization**" or **SEO**, is generally beneficial. However, it doesn't always result in more accurate tagging and better content, but sometimes worse. For many sites dependent on advertising, Google's ranking is far more vital than quality of the information offered.

Various tricks can be used to artificially inflate rankings. Many of these sites are "**aggregators**", mashing together information gained by "**content scraping**". Since Wikipedia, for instance, usually ranks high in Google results, its information is copied and pasted to other sites in hopes of boosting their own scores. Sensational stories attract attention, and so inaccu-

Continued on back



**Southwest
Cyberport**

New Mexico's Leading Local Full-Service Internet Provider Since 1994

swcp.com | help@swcp.com

5021 Indian School NE, Suite 600, Albuquerque, NM 87110-8910 USA

505-232-7992

Continued from front

rate and even misleading headlines and content tend to proliferate. Plus, there are often more pages to click through to get to what you want. Google says they are now working on these annoying results of their own success, but it may not be easy or simple.

For years, there have been accusations that Google actively censors its page results, that certain sites are consistently downplayed or omitted altogether. Unfortunately, such fears aren't groundless. To gain access to the vast Chinese market, the company apparently agreed to cooperate with government censors including content filtering. Google continued until they were hacked by the Chinese and ended cooperation. But how much, if any, other filtering still goes on elsewhere or what criteria govern it is anyone's guess.

However, the story points out another good reason to look beyond Google: **different results**. All search engines depend on their own rules, called **algorithms**, that tell them how to search, what elements to look for, and how to build the index. These results may be radically different depending on what the programmers decided was important.

For instance, Microsoft's **Bing**, www.bing.com, bills itself as a "decision engine"; which gives greater weight to activities than to passive facts. So a list generated by Bing might be much more action-oriented than one looking up the same term by Google.

Specialized search engines are becoming more frequent. They allow you to more precisely center your search on what you need. Often these tools do not stand alone, but are associated with large content portals. The **Internet Movie Database**, www.imdb.com, has one specifically related to visual entertainment topics, directors, actors, and so forth. **Goodreads**, www.goodreads.com, a crowd-sourced review site, has one all about books and authors. And while the search functions at **Amazon.com**, www.amazon.com, are limited to the bookstore, its products and affiliates, that's still a gigantic pile of possibilities.

As Bing moves to define action steps, other tools work to give more complete information. Some, like **Ask.com**, www.ask.com, simply mash together the outputs from a number of sites. Much more ambitious is **WolframAlpha**, www.wolframalpha.com, which calls itself a "computational knowledge engine." It intends to solve problems, do math, and give answers aided by experts and a community of users. While very promising, computing's ancient challenge of phrasing questions that can be productively answered by machines still needs a bit of work.

Semantic searching - that is, based on understanding the meaning of the terms involved - is the end goal. However, since semantics requires artificial intelligence, it may be presently unattainable. But the need is obvious, if, for example, you've ever tried to find a piece of artwork just from a description.

For visual images, search engines do little more than the "find" function in your word processor. They home in on words nearest the art on the page, the file's name, or terms hidden in the code. If you're familiar with the name of the artist or the title of the work, this means you usually find it easily. But if you can't identify a painting apart from the surrounding text; you're stuck: even dedicated search engines such as the one at the popular photo-sharing site **Flickr**, www.flickr.com, can only identify large area of color. For all else, it still relies on human-placed tags.

Search limitations

Like our minds, search functions work by association, but are unable to make the intuitive leaps that we do effortlessly. What if all you knew was that you wanted a certain "Renaissance painting of a dark-haired, smiling Italian woman"? You might instantly think of the Mona Lisa, the most famous picture ever; but it would not be so easy for a computer.

While IBM's **Watson** supercomputer was able to beat humans at a game of *Jeopardy*, that data-crunching power is quite unique. It may be some time yet before a common web application has similar abilities. Yet the truly critical constraints remain not those of machines but of *how* people use them.

After all, we humans have stubborn opinions, hot buttons, even blind spots. We build world-views that are can be intellectual comfort zones or cages. They can invisibly screen from our awareness what we don't want to see. A term coined recently by thinker Eli Pariser, "**filter bubble**", www.thefilterbubble.com, warns of the understated but widespread hazards from too much personalization of the Internet.

The "filter bubble" means that while it may be very pleasant only experiencing online content that's tailored to your tastes, it's also limiting. A fully customized Internet would be like a shimmering hall of mirrors. Your view of reality could become so distorted as to result in unfortunate choices. And you might easily miss out on many invaluable opportunities for happy surprises and real growth as a human being.

Filter bubbles are already firmly in place, growing ever stronger with increased customization. Even now, other people may receive different lists of results for the very same terms from Google than you do, depending on their profiles and history. Amazon is even more aggressive with a slew of choice recommendations the moment you return. Even news sites are succumbing, which can only increase potentially dangerous political polarization.

So it pays to climb out of your Internet rut once in a while and look around. Every now and then, check out alternative sites to those you rely on each and every day, including search engines. While Google is an awesome, useful tool, you may find unexpected delights simply by stepping beyond the familiar.

