



Entrepreneur Incubator

Reference Article

Google and what it does

August 2008

Have you “googled” today?

If so, you’ve seen the future, and you are part of it, even if you don’t realize it just yet. In fact, there’s a lot more going on at Google than just searching. I’m going to talk about what Google does, knows — as well as what you should do and know to get the most out of it.

True enough, Google is best known for its powerful search engine, which draws upon a massive database of web pages, videos, images, stock quotes, phone numbers, addresses, and other chunks of media and data to provide relevant information — all in the time it takes to strike a key.

The name of the world’s largest search engine has become synonymous with searching the Internet. It’s even become a verb in our everyday language (even though Google doesn’t like us using it that way): we “google” car reviews, our childhood friend’s current address, and, in flights of fancy and grandeur, ourselves.

It is easy to forget that we’re not alone when we’re searching. Everything we do within the Google environment (and some things we do outside of it) are recorded, stored, and analyzed. Once you know that a company has that much data available, it doesn’t take long to realize that all this information could be used for far more than just a pleasant search experience.

To fully understand Google’s potential power, sit back for a moment and consider how Google works in a bit more detail.

How Google Works

To the casual searcher, it isn’t immediately apparent how Google searches the Internet so rapidly and with such good results. The secret is that when you type in a query they aren’t searching the Internet at all; instead, Google is searching a database that is constantly growing and updating, 24 hours a day, 7 days a week.

All this happens without any human intervention. Tiny programs called “spiders” or “crawlers” independently follow links from one page to another, the way a spider follows the threads of its web. As the spiders crawl across web pages, they collect information about every sentence, image, phone number and anything else they might encounter. They scan each page they visit, index keywords and note links to and from the page. Then this information is stored in a massive database. The same links and pages are crawled again and again, ensuring that Google’s information stays current.

Google ranks pages using the information its spiders collect. Included in this ranking process are the following:

keyword frequency and location (more keywords in more prominent locations are better);



Entrepreneur Incubator Reference Article

the age of the page (the more well-established the page is, the better); the number of pages linking to a page (the more links the better). There are other factors Google considers as well, but these are kept secret to hinder those who would try to game the system and score high ranks without actually providing useful information. In addition, if you overdo it and simply fill your page with redundant information, lists of keywords, and junk links, Google will penalize you or may even drop your site from the database. Clearly, a fine balance must be struck. That balance of factors on the page is as much art as science.

The better a page scores according to these criteria, the higher it ranks in the search results. The highest-ranked pages appear on the first page of a Google search; since most Google users never venture past that first page, those high ranked pages get a disproportionate amount of Internet traffic. All this to say that, when you run a search on Google, it can respond so quickly because it is not trying to search the entire Internet at that moment; it is consulting its highly organized and prioritized database. But even more important than creating lightning-fast searches, this means that Google has a vast repository of data about what is on the Internet and, more importantly, who is using what, when and for what purpose.

What Does Google Know?

Google "knows" a lot, actually. Google keeps track of searches, and even keeps a file on your particular searches based on your IP address or Google login. That gives them a little window into your mind, revealing what interests you, worries you, excites you and frightens you. This alone is powerful marketing information. If you could tap into that database, you would be able to construct detailed profiles about individuals — their interests, buying habits, health concerns, family issues, and more. You can uncover signs suggesting whether a company is succeeding or failing, whether it is considering a merger or acquisition, and what product lines it may be expanding into. You can track historical trends in elections, economics, health care, and any number of other areas that have significant social, financial, and political value.

Moreover, if you, knowingly or unknowingly, make use of other Google services, you are providing still more information. They scan Gmail, gaining the same information from your correspondence as your searches, as well as whom your friends and collaborators are. If you share documents through Google docs, they know who you work with, and on what projects. Google Checkout adds data about your purchasing patterns, your spending habits, and your budget. There is also information that you transmit to Google without even being aware that you are doing so. Google collects, as do other Internet services, your IP address and possibly your MAC address. Your IP address tells them roughly where you are in the world geographically, and your MAC address is a signature unique to your machine; this allows them to track whether you always use the same machine, how many machines you use, and so on. So, not only does Google know what you are searching for and whom you are communicating with, the



Entrepreneur Incubator Reference Article

company also knows where you are and which machine or machines you are using.

Of course, let's take a step back and admit that almost any Internet site has access to some or all of this information. For example, IP addresses are regularly collected to defend against hackers and denial of service attacks. Information about where you go on someone's site, when you arrive and leave, and where you come from and go to, are all standard web statistics available to even the smallest personal site.

What makes Google different is its sheer size and diversity of services, allowing it to collect a greater variety of information from many more people.

When your stash of information goes from hundreds of pieces of data to billions, you have insight that no one else does. What makes Google truly powerful is that it can observe people in more contexts than anyone else. Google is watching you even when you aren't on your computer. Google Maps provides high-resolution pictures of most of the United States, as well as some other areas of the world. It is likely that your home can be seen on Google Maps, and, if you happened to be watering the yard when the satellite flew past, you may also be able to see yourself in your rattiest pair of shorts!

What Google Can Do

There are some fairly obvious uses for Google's database, AdWords perhaps being the most popular and visible one at the moment.

Each time you run a search on Google, the first few listings at the top of the page, and the listings on the right-hand side are "sponsored links," paid advertisements bought by people who believe that searchers who run a search like yours would be interested in their goods and services. This technique has spread to other web pages, which explains the "Ads by Google" moniker you see delivering relevant ads for the pages of other websites.

Clearly, selling this advertising space is an excellent source of revenue for Google. In 2007, AdWords brought Google over \$16 billion in revenues, making it Google's largest source of income so far. In comparison, the Google data store is a gold mine that has barely been tapped.

So far, the Google features that we have seen have been focused primarily on reacting to current market trends. People become interested in something — the new version of the X-Box, finding an electrician online, Super Bowl memorabilia, etc. — and Google is positioned to help people connect with their potential customers.

However, with enough information, a company like Google can do more than simply react to the present with lightning speed. It can also see the future, or even create the future. Before you think this is just a paranoid science-fiction daydream, give the issue some consideration.



Entrepreneur Incubator Reference Article

Let's take a simple case. Suppose you set up a program to note searches that fail to turn up any highly ranked pages—failed searches that do not provide any really useful information to the searcher. The program that tracks these failures notes what the searcher wanted, and puts those failed searches in categories. Looking through those tallies, what if Google notices that there are a large number of failed searches that all have to do with finding a do-it-yourself superstore in Thermopolis, WY. A little more looking about, and I have a list of places where there are many failed DIY superstore locator searches. That list would be incredibly valuable to a company that franchises DIY stores (not naming any names, but think of those big orange warehouses in every suburb: wouldn't they like to know?).

They would not only know that there was no DIY store there, but that a specific number of people were looking for such a store. And they would probably have some idea what they hoped to buy there: appliances, building materials, gas grills, etc. This is painless market research, neatly sorted and analyzed by Google automatically.

In the same way this information could be used to identify potential buyers for a particular product, track criminals engaged in child pornography, and catch potential thieves casing a home or store. It can help developers choose sites for new homes and stores, warn you of severe weather in your area, locate registered voters and learn about their habits and interests, and help the IRS find people who cheat on their taxes. There could also be mistakes of interpretation: a bunch of searches about cancer from a small town might mean a lot of cancer patients, but it could just as well mean there is a medical school there. As you can see, we may applaud some of the uses of this information, and decry others. But they are all possible, and all use the same basic techniques. You can find nearly anything you want, you just have to know how to look for it.

What Does This Google Data Mean For You As An Individual?

Although Google has resisted providing easy access to their terms, in July 2008 they finally added a link to their privacy policies at the bottom of each page, next to the copyright statement.

As a person whose information is constantly being collected, reading this notice may all be a bit disturbing.

Google has pages upon pages laying out their privacy policies, with separate pages to delineate the particulars for each service. Wading through these is a daunting task, and doing so may not answer all your questions to your satisfaction. There have been consistent complaints about the fact that Google's privacy policies are vague, something which Google admits. Writing clear policies for so many different products, all of which are constantly advancing and evolving, is a formidable challenge.

The key issues for a user are probably these:

Google stores non-anonymous data in their logs for 18 months before stripping



Entrepreneur Incubator Reference Article

your identifying information from the activities you carried out at Google. It is not possible to remove yourself from Google; the nature of search engines is that they draw from so many highly interconnected sources that, practically speaking, it would not be possible to satisfy requests for removal of information. This means that a great deal of personal information is available, and remains available long enough that Google and its clients could build extensive profiles of everyone who uses any of Google's products.

What Does This Mean For You As An Entrepreneur?

Ironically, the very things that make us uneasy as individuals may excite us as businesspeople, and fill us with ideas to grow our market and expand our product line. We can not only make the best use of current opportunities, but see new ones on the horizon, and even shape the future to create opportunities tailor-made to our plans and goals.

As a businessperson you can tap into this information store by using Google's massive traffic map and ability to send targeted traffic to your sites. Not only can this generate business and revenues for you now, it can provide you with vital information for building your business in the future.

You can access this information by using Google's paid services for advertising, such as AdWords and AdSense to drive traffic with relevant interests to your site, though you will need a substantial advertising budget if you take this route. Popular keywords often require you to commit large sums in order to ensure that your advertisements will be seen frequently and at the top of the list. A better alternative is achieving a high "natural" or "organic" ranking. This refers to using strategies that result in your pages being highly ranked in Google searches without you explicitly paying to get those coveted positions. The top search results on the first page of a Google search are far more likely to be visited by Google users than lower-ranked pages, and can even outperform paid ads.

By fully understanding what it is that Google does, how it does it, and what information it provides you, you can combine organic and paid services to maximize your ranking, visibility and traffic. And together, that means maximizing your revenues, too.

As I discussed above, it's not as simple as using lots of keywords and links — these need to be embedded in legitimate content and strategically placed for the highest impact. You also need to track which topics are getting the most attention from web users, know how to translate the different types of information Google provides into insights for your personal use, and how to integrate multiple services for the best impact. All this is simple once you have the right information and the right tools. Information is the coin of the realm.

Hamlet Batista is President of NEMedia S.A, a provider of [SEO automation software](#) that helps entrepreneurs and small businesses increase the quality of



Entrepreneur Incubator Reference Article

their natural search traffic while focusing on what they do best. Hamlet's blog, [Hamlet Batista dot Com](#), explores the most advanced SEO research, as well as strategies and tactics that can give you an important edge over your competitors.