To be on top of the search list is to be on top of world.

Hidden Mysteries of SEO

There is an arcane art that keeps webmasters up at night anxiously tweaking their sites. Billions of dollars are spent on it, and uncounted thousands of consultants claim mastery, but none can fully explain just how it works. Both honest merchants and spammers study this art carefully, for its practices run from "white" and most benign to "black" and unethical.

This net wizardry is known as **SEO** – **Search Engine Optimization**. Ordinary users may never have heard of it, but SEO is central to every search they undertake and each webpage they visit. Recent concerns about "fake news" make it even more important.

Location, location, location

Optimizing pages for search engines is not simple: it's a collection of strategies, methods, and tricks to increase traffic to websites by getting the site listed in first place in online search results. Search is vital because the web is so vast that it desperately needs an index. Not only that, it needs a catalog that ranks listings in terms of page quality and usefulness.

Search engines are the means of generating that catalog. Today, there are **4.59 billion pages indexed** by Google and Bing. Google is dominant, **accounting** for 76% of all desktop searches and 94% from phones. Bing, Baidu, and Yahoo each have 8% or less.

But people are funny. Though a search may pull up millions of related pages, most people pick something off the top. 75% **never get beyond** the first page of listings, and 80% ignore sponsored results.

So the top of the first page of a Google listing is the most valuable and contested real estate on the web today. SEO is all about landing their websites in those coveted positions, and competition for those few precious slots is enormous and constantly growing.

One reason that users go for the initial links on the page is the **presumption of quality**. They assume that those pages will give them the information they seek without annoying distractions. Google and the others go to enormous lengths to try to do just that.

It's not just that there's nothing as pathetic as a blog nobody reads. There's **big money** involved, too – retail digital **advertisers** spent over $15 billion last year alone. Yet that's **dwarfed** by the $65.26 billion

spent on SEO in the US in same period. This shows not only how essential online marketing has become to retailers, but what a costly mystery good SEO is.

One reason is that print and traditional advertising is as dead as print media, and TV ads are easy to ignore, so the real ad game is online. Especially since half of all visitors to a search results page cannot recognize inserted paid ads from unsponsored listings.

Behind the screen

In the beginning, the closest thing to an index was a human-made list of links. Yahoo began that way; and until it closed in March, **DMOZ**, the Open Directory Project, had been hand-indexing for almost 20 years. Yet in all that time, it listed just 100,000 pages.

Automation was obviously needed. Searching is easy to do in theory as computer search functions simply compare one string of bits to others, looking for matches. The really tricky part is judging how helpful the match is, and that's where SEO comes in.

Webmasters quickly saw how important it was to get a high ranking, and tried various means to inflate them. Some aided users, like **organic links**, unasked-for inbound links from other pages to that page, which showed how useful other webmasters found it. Others were outright frauds, like hiding content that only **spiders** (the programs that crawl the web scanning sites) would see, but not readers.

Organic linking is still the key to SEO. 70% of the links users click on are organic. A **study** of Google's pages show that 18% of links are organic in the first rank listing, 10% in the second, and 7% in the third. For Bing, it's about half that. But the more links a web publisher can get coming in, the better the ranking.

Of course, it gets much more complicated than that. It is hard work to legitmately increase traffic by generating great content and links to other good pages, and much easier to try to deceive the search engines.

That, however, if detected, can result in getting a site banned or deprecated in the listings. But still the black hats try to pull it off, seeking to fool naive users or use some new trick for short-term gains. For Google and the others know most of the tricks already. Some they condemn publicly to discourage use, others are kept secret, especially if prevention is hard.

Ranking is done by complex mathematical formulas called **algorithms**. Exactly how they work is a closely-

held secret, but Google constantly develops code-named improvements to them to block the latest black hat scams. For example, **Panda** in 2011 was introduced to lower rankings of **content farms**, which generate lots of stories with little real meat.

But in all their improvements, the criterion of **objective truthfulness** never came up before this year. Google and others are now trying to evaluate pages for honesty, and are finding it very difficult. As with every change in how they judge pages, this has some legitimate websites – in this case, various political advocacy sites – finding their rankings dropping drastically, darkly mutter about **media conspiracies**.

Google withholds the exact formulas for white hats, too. But they are very generous on detailing tweaks webmasters can use to improve their sites in general.

Most of these are very technical, involving site structure, page-naming conventions, and generally making things easier for their skulking spiders. But they constantly stress that quality content and links are the most effective way to increase those rankings.

And of course, they provide loads of data to help webmasters, and particularly advertisers. Google came under fire when it was **discovered** that ads were being sold to neo-Nazis and jihadists on YouTube. Not only that but Google's **AdWord** generator easily enabled anybody to market directly to them.

Seeker as product

What advertisers really want to know is the **click-through rate** (CTR). That's the number of times a visitor will click on a link in a listing or page, divided by the number of times the listing or page has been seen. It's the defining metric of digital advertising, as the people interested enough to click may indicate how well the ad works. But like all measures of advertising success, it's hard to know how really useful it is.

Google's avowed aim is to catalog all the information online and make it available to everyone. But, as a commercial enterprise, it's not doing that out of goodwill. Google, Yahoo, Bing and others, are essentially **advertising platforms**. They want to find out what interests users in order to sell them stuff.

As users are presented with these glittering bazaars, they provide search engines, and the marketers who employ them for advertising, with vast amounts of data. The search engine then mirrors that data back at them, giving them more similar choices.

Users thus find themselves lost in an infinite hall of mirrors reflecting their likes and dislikes. Is it any wonder that they tend to mistake this for the real world? Much of the inflexibility and one-sidedness afflicting civil discourse today can be blamed on this effect. This overt pandering not only directs search engines, but all sites where users are the product, like Facebook, Instagram, Twitter and on and on.

What is worse, studies have shown that **manipulating search engine results** can change users' behavior, even **voting patterns**. Even simple things like using **autocomplete** to finish search terms can skew outcomes, simply by which words come up. And it appears to **move results** to more extreme positions, as rabid partisans seem to be the most active online.

Google was **forced** to remove antisemitic results once this was found, but the problem is getting worse for political groups may have been exploiting this effect for years. Tweaking a page's actual ranking has an even greater effect, so there may be **something** to liberal fears of conspiracy after all.

Looking beyond Google

Google is the 600-pound gorilla of search, but it's far from being the only search engine around. There are **hundreds** out there, including many specialized engines for law, medicine, and even video games.

There are engines that specialize in finding people or TV shows and torrents, plus map generators, job lists, real estate and so on. And many large sites, such as the **Internet Movie Database** and **Facebook** have their own internal and external search engines that work off their own specialized databases.

Search engines come in a variety of languages, and there are even "**metasearch engines**" that aggregate searches from a variety of sources. Many, however, are based on Google, Bing, Yahoo, and even Ask.com.

One kind of search where much work is needed, though, is **searching for images**. **Google** and Yahoo offer image searching, but since it's all based on keyword descriptions, it's still far from accurate.

Why use anything *other* than Google? Since all search engines depend upon their own proprietary algorithms, not all will come up with the same results to identical search terms. Plus, commercial engines are greedy for user data, which they sell to advertisers. If a user desires anonymity, **DuckDuckGo** can search using **various sources** while preserving privacy.

Finding needed information is an essential skill, and there are **ways to do it better**. And sometimes, it pays to look beyond the first results. There may be unsuspected hidden gems just waiting to be found.