Search Engine Design – understanding how algorithms behind search engines are established

An article written by
Koulutus- and Konsultointipalvelu KK Mediat

Web: http://www.2kmediat.com/kkmediat/eng/
Last updated: 09.08.2004

Originally published at SEOGuy.com

Introduction to the world of information searching

As Search engines success, many users have noticed that they are overloaded with information. Despite the common errata, the amount of information available has not increased. The information overload is due to fact that search engines have brought data more accessible. And as companies, organisations etc. transfer their archives to public electronic format, the problem will only get worse.

This may lead to severe problems, since many good, very relevant sites get buried somewhere in result. Users have no means to tell whether or not the data they received is relevant, and in many cases they notice the existence of irrelevant results. A large part of SEO and SEM is making content more accessible on search engines. The purpose of this article series is to describe how and why search engines work the way they do and how to implement this information on SEO / SEM strategies.

Directory versus search engine

Directories are hierarchical index lists of sites; they list sites by topic. They are widely used and in many cases offer an extremely great source of information. However, they have few problems:

- Hierarchies are very vulnerable. Data and its classifications change constantly. This also leads to changes in hierarchy. A good example of this is DMOZ, world's largest directory. Several subcategories are created, removed or deleted each day.

- Most directories rely on human intelligence and are manually edited. They can never compete with search engines in amount of information. However, quantity is never as important as quality.

Unlike directories, which show the hierarchy of data, search engines have only one purpose: tell user where to find the most relevant data. Search engines rely to mere computing power instead of human intelligence that directories offer. Robots crawl web pages, analyze and split them to smaller pieces. As user enters search phrase, search engine returns pages containing searched words sorted by relevance. This relevance is

determined by something called algorithm.

In theory search engines sound picture perfect solution, but sorting out only the relevant results is something that no search engine has yet managed to do. Although there is clear evolution in algorithms search engines use, the journey ahead is long and rocky.

The golden egg of Search Engine Design

According to several researches over 70% of computer users experience that finding relevant data fast and easily would be very important to them. No wonder that many companies are racing in this area. Who possesses the best technology, is likely to become extremely rich and powerful.

The bad news is that over one third of users experience the opposite. For search engines and developers working with them, this is big issue. The million dollar question is" how to help user to find relevant and accurate information fast and easily?" This is something where search engine design plays a major role. If you wonder what search engine design is and what is has to do with your SEO / SEM campaign, read on.

To find information people use various strategies. Some prefer to use search engines, some directories, others mail a friend or make a lucky guess. One of the most important issues that affect how search engines work and what direction the algorithm is going, is answering the questions like "How users seek information?" or "What kind of strategies they use?". These are the basic questions that Search Engine Design tries to answer and react. Search Engine Design is understanding how people seek information and modifying search engines so that they match expectations as well as possible. Everything else is only the cause.

Understanding this puts traditional search engine optimization and marketing to new light. SEO / SEM tries to outsmart how search engines algorithms are build, in other words find mathematical formulas behind Search Engine Design. To crack the algorithm and success well in search engines, it is crucial to answer basic Search engine design questions like:

- what is my target audience
- what kind of searchers they are
- how they seek information
- what kind of information are they seeking
- what keywords do they use when seeking information

The old rule "make pages for users, not for search engines" gets a brand new glow. The search engines rank pages / sites by how well they match to Search Engine Design issues. Issues that define Search Engine Design are pretty much defined by user. A SEO can manipulate backlinks, keywords etc. to very large extend, but finding the right balance is difficult.

The next chapters try to answer the basic questions of SED and highlight some of the issues that might be useful when building a SEO strategy.

Identifying target audience and their search preferences

Users can be divided to two main branches based on how they search information. Link oriented searchers prefer reading articles and following links on pages. Search oriented users on prefer typing search phrase to box and hitting go!-button. Tendency to either kind of behaviour is a combination of multiple factors: lifestyle, hobbies, work and family relationships etc. In general, technically advanced users and experts prefer search oriented approach. If you think this has nothing to with SEO / SEM, prepare for surprise.

About 20 percentages of users are link oriented. Considering SEO / SEM this group may sound somewhat uninteresting, since they don't tend to use search engines to find information and services. However, identifying this group and their interests is important, since they play a major role in entity.

Generally link oriented users and subjects are relatively easy to notice. Or have you not ever wondered why some pages receive loads of free links? Link oriented users like good content: they produce good content and they like to link to good content. And as you know, links are important. And if those links come from quality (authority) sites, that is even better. As an added benefit, most link oriented users are potential "regulars" (people who come time after time to your site). So getting in touch with link oriented searchers is pretty important.

Search oriented users are totally different case. They like to do a hit and run mission. About 70% of users belong to this group and their level of expertise varies a lot. Their behaviour is pretty much controlled by search engines and results they give. This is both a benefit and a risk. In eyes of SEO / SEM, search oriented users are the easy way to riches.

Most search oriented users are either
a) Relatively young (below age 30) or
b) Experts or
c) Technical minded

Most of the search oriented users do basic searches:

- Search phrases contain one or two words. The general thumb rule is "shorter the better", which often leads to very wide set of results and not finding what they are after. Users might type search word **ringtones**, but in most cases they are after something more specific such as **Anastacia ringtones.** Concentrating on longer search terms gives benefit both to searcher, Search engine designer and SEO, because targeting for relevant results is easier.

- Used search terms are generally of basic, speech form. This means that when searching an activity, people use forms of **doing**, not **to do** format (painting / to paint, skiing / to ski etc). Also identifying and targeting the right key terms is important. Or what do you think is the most popular search term for possible client to look after SEO services: SEO, Search Engine optimization, SEM, Search engine marketing or something totally different?

Largest search oriented group are young users. They belong to internet-generation. They use search engines to provide mainly entertainment.

Experts and technical minded users are their own breed within search oriented users. They have extremely good skills of finding right terms with correct nuances, forming complex searches and finding relevant results. Targeting for them requires good knowledge of special expertise / subject related glossary.

How do search engines take user types and their preferences into account? Do a simple test and do various kinds of searches. Be a link oriented searcher (search something informative such as ”CSS learning resources”), be a simple search oriented user (search for some entertainment with simple 1-2 word searches like ”usher”) and finally be a technically oriented user (search for ”hp DeskJet 990c cartridge”). Take a look at keyword density, backlinks, related sites, site structure / layout etc. Do a simple analysis on the site and topic you are watching - notice the differences? So do search engines.

When launching a SEO / SEM campaign, it is vital to recognize what user types your sites will attract. This defines what types of keyword combinations and information you should provide in your site. The decisions you make have an extensive affect how you will rank.

Re-defining the search process

Search process can be split to four stages:

1. formatting search phrase
2. performing the search for search phrase
3. validating search engine results
4. Reformatting search if necessary.

SEO and SEM have one target: to achieve good rankings for specific key phrases in specific search engines. The focus is in part 3 (validating search engine results). Although there's nothing wrong with this approach, it is relatively narrow. When you start considering how people make up their mind which key phrases and result(s) to choose, you'll begin to ask questions like:

- how users choose and format the search phrases
- how users choose a specific result on the page
- how to improve ROI

Formatting search phrase

Formatting search phrase is most critical and difficult phase when performing the search. First the user must come up with appropriate search words and then format a good search phrase. Many usability tests show that this is the hardest part for regular users. One of the most common reasons for giving up is that users don't know what terms to use or how they are called in the web.

Unlike machines (like search engines), human memory works with associations. Humans don't need to know exact details to perform operations well. However, for computers and computer applications like search engines this is critical. A good example of this is search word order. For human ”man and child” and ”child and man” bring up pretty similar image, for computer application the result is totally different. This is one of the issues search engine algorithms try to handle: make search term order as transparent and natural as possible.

One problem for search queries is that human memory is content specific. Human remembers specific terms only in conjunction with specific issues. If the term comes up in another relation, the association is not created. Computer however compares single words and it has no such limitations (or benefits). For this reason we humans are very poor at

coming up with synonyms and innovative search phrase reformatting.

The more detailed the search, the better results search engines can give. Shorter search phrases on the other hand bring up more associations, and it is more than likely that the number of irrelevant results increases. A bit over 40 percentages of search engine queries are single-word search phrases. It's been a clear trend that users are slowly learning to write more specific searches, but the process in relatively slow. And every generation needs to learn from the start.

One of most interesting studies is that the length of search boxes is related to length of search phrase. This is most likely the reason why almost all search engines have introduced new layouts with longer search boxes.

Performing the search

Performing the search is at the same the biggest mystery, but also the most transparent action. Basically hitting the Go-button is all that is required to perform the search. More complexly, it is a bridge between user (searcher) and application (search engine).

Operations happening behind the scenes (database queries, algorithms) are the cause of Search Engine design. Search engine design dictates how various ranking factors are valued and more importantly, why are they valued that why. And when we understand that, we can get a pretty clear image how algorithms work.

Validating results

Validating results is relatively easy for most users, since it's a natural response a set of stimulations. The response can be activated by a single word, sentence, colour, phrase etc. factor. A large part of successful SEO / SEM is offering users the right kind of stimulation. When they see a certain page title or sentence, they automatically know they have found what they were looking for. Getting to page one is of no use, unless you get noticed.

A somewhat classic psychological division of how people make choices is:

- The intensity of stimulation; for example brightness of light, loudness of sound etc.
- The importance issue

Take for example search term *PHP hosting*. If you are on the move with some serious hosting needs, key phrases like *99,9% uptime* and *24/7 support* are more likely to catch your eye than phrase *affordable budget hosting*. And of course the factors go vice versa, if you want something cheap words like affordable and budget are pure candy to you.

Intensity of stimulation in search results combines of multiple factors. One of the most important aspects is the fact that users look specific information from specific positions on screen. If the data does not match to what they are looking, they don't never even notice the issue. This leads to presenting information to people. Getting the information on the right spot on the screen and offering them the right words.

A good example of this is Google results page:

- Only10 first results are shown initially. Only 4 or 5 first results are shown without scrolling down the page. This puts additional visibility to first results on pages.

- Titles are the dominant factor when viewing results. They are represented in a bit larger font with bright blue colour. This puts them clearly on the map.
- Page extracts show sentences that are thought to be relevant to search. Search terms are bolded in extracts to put additional stimulation to correct words. These extracts help user to form expectations by which he/she makes the choice.

Nielsen marketing has studies on how people read online text. It comes to no surprise that long text entities are rarely read in whole; almost 80 percentages of users skip more or less text. One of the possible reasons for this is so called priming phenomena: certain words or word beginnings activate user; in our interest these are search terms. The problem is that human user is easy to distract if there are many similar words. Especially long words and/or sentences and foreign based words are the ideal for distraction.

If the user is not satisfied with results, he either gives up or reformats the search phrase. This is something many do very poorly and usually end up changing word order instead of trying synonyms or adding new targeting search phrases. It seems that for regular user observing the issue from another perspective is extremely difficult.

A summary

What have we learned from the above article? The basic punch line is that search engines form their algorithms and results based on how users search information and how to provide them information that is relevant (in that specific order). If users change the way they seek information or webmasters change the way they produce content, search engines reply with algorithm changes.

In order to do long term, successful SEO / SEM campaigns, broadening the vision to include Search Engine Design is mandatory. In other words this means “make pages for user, not for search engines”. This doesn’t however mean that the end of SEO/SEM is near. A lot of SEO has to do with accessibility (such us appropriate link and site-structure, simple filenames, appropriate headers) and traditional marketing. These are a major factor and guarantee that a good SEO is never out of work.