Search Engine Optimization -
From Automatic Repetitive Steps To Subtle Site Development

Robertas Narkevičius [a]

Vilnius Business College, Kalvariju street 125, Vilnius, Lithuania

Abstract. Nowadays, position of website in our search results is determined by using hundreds of factors designed to provide end-users with the helpful and accurate search results. These dynamic factors have been changing every year or rather every month. This work is devoted to the analysis of search engine optimization problems from the end-user perspective.

Introduction

Websites play a very important role in the processes of information spreading – from domain to final users. Efficiency of spreading is a nonlinear function belonging on many cross overloaded parameters. According to the empiric experience such problem could be formulated as follows: many websites pay a lot of money to increase the web traffic, also they spent lots of time for traffic optimization which results in increasing amount of users and finally disappearing from Google.

Google started to use more complicated search methods in comparison to previous ones (well known in the previous decade), such as "Penguin filter" and "Panda filter". Searching procedures have been enriched nowadays with parameters related to the visitor's behaviour. Nowadays, position of website in our search results is determined by using hundreds of factors designed to provide end-users with the helpful and accurate search results. These dynamic factors have been changing every year or rather every month.

This work is devoted to the analysis of search engine optimization problems from the end-user perspective.

1. Solution to the effective search problem

Many previous search rules are to be reconsidered according to the requirements of nowadays. Search engine optimization (SEO) is the process of affecting the visibility of a website as a search result in the search engine. Some web-researchers claim, that SEO is disappearing. However, there is also evident that SEO has become more sophisticated, more unpredictable. Nowadays, the main paradigm of Google according to the search processes is formulated as follows: **"Web site is for people, not for search engines"**.

Table 1 represents Google quality guidelines expressed as the basic principles according to Ref. [1]. There are seve-ral questions to be considered as a "rule of thumb", whether you would feel comfortable explaining what you have done to a website that competes with you or to a Google employee. Another useful question is if this helps my users or "would I do this if search engines did not exist?"

Here, the main task is to create unique, useful content for people. Interesting sites will increase their recognition on their own. If you show useful content, then the SEO becomes popular and important. According to the SEO later you can make some changes you want. It is also important to work on useful keywords. One more important aspect is to be able to create unique content on the site, to avoid duplicating pages. There are sites which have unique content, but it is repeated and dominant. When the content is small, or too small it may become irrelevant for the users.

It is difficult to say, if the SEO will work efficiently without the unique content. On the other hand, without knowing the rules, this unique page might not be able to achieve the task wanted. Here, the SEO is based on creative approach rather than repetitive automatic rules. Of course, site owners should be aware of tools, which might help to understand web site situation better.

Many other SEO features exist, and some of them can affect web page greatly according to Ref. [2]:

i)   internal Links (Inner URLs, site structure);
ii)   external Links (Outbound URLs);
iii)   anchor Text;
iv)   title and Meta Description;
v)   on-Page Factors, H1, H2, H3 elements on a page;
vi)   duplicate content and redirection;
vii)   robots.txt;
viii)   HTTP Status Codes;
ix)   domain name.

[a]email: robertas@kolegija.lt

	Table 1. Google quality guidelines.
1.	Make pages primarily for users, not for search engines.
2.	Do not deceive your users.
3.	Avoid tricks intending to improve search engine rankings.
4.	Think about what makes your website unique, valuable or engaging. Make your website stand out from the others in your field.

2. Rules

2.1. External Links

Most SEO experts agree, that external links are the most important in page evaluation process. External links are the most important source of ranking power. If another website links to your own website, this is considered to be an external link to your site. Similarly, if you link out to another website, this is also considered as an external link. Do we need to put external links anywhere? What sites are the most important?

Thus, SEO has become more complicated for links of such type. Today, the major search engines use many metrics to determine the value of external links. Analyzing the suggestions of experts, the following important criteria can be extrapolated as presented below:

i) trustworthiness of the linking domain;
ii) popularity of the linking page;
iii) relevancy of the content between the source page and the target page;
iv) anchor text used in the link;
v) amount of links to the same page on the source page;
vi) amount of domains that link to the target page;
vii) amount of variations that are used as anchor text to link to the target page;
viii) ownership relationship between the source and target domains;
ix) PageRank.

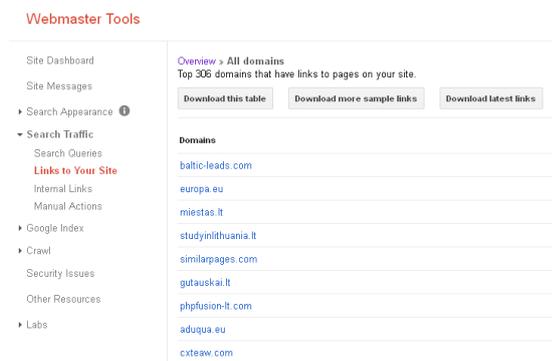Many specialists of SEO agree that the importance of Google value - PageRank decreases time after time.

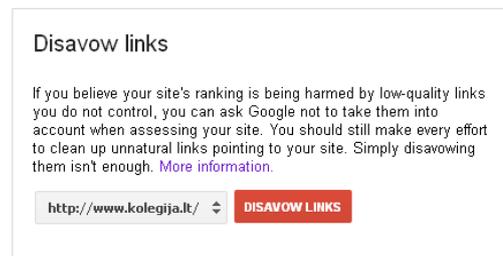Fig. 1. "Webmaster tools" environment - downloading links.

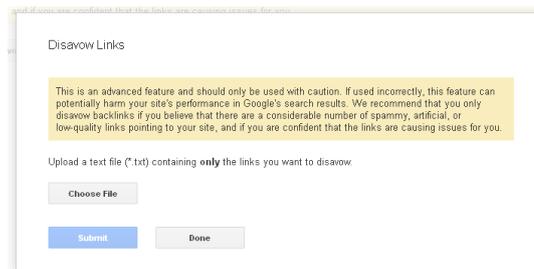Fig. 2. Disavow tool environment- choosing domain.

Fig. 3. Disavow tool environment - setting file.

Google tries to interpret quality of link. If some link created by owners is placed on spam sites (or sites, which re-sells links; or some links are repeated on outbound site many times; or too frequently comparing with the other sites, unnaturally) - this link can decrease your page ranking. In some circumstances, the incoming links can affect Google's opinion of a page or site negatively. Then web site owners can use the special link of so called "disavow tool" type. In other words, owners can ask (request) Google service not to take into account the certain external links - by assessing owner's website. As a result three-step operations must be done.

Firstly, in order to do that, a website owner is a first person must register and verify ownership of another tool "Google webmaster tool" [3], see Fig. 1.

Secondly, website owner needs to download link table and to run "disavow tool" and upload the needed link list. Fig. 2 represents mentioned routine constructing using Ref. [4].

Thirdly, the owner can approve a list of bad, spamy, not needed link, see Fig. 3.

2.2. Internal Links

The Internal Links page lists are the set of pages on your site that have incoming links from the other internal pages. Hundreds of thousands of sites make the critical mistake of hiding or burying their main link navigation so that search engines cannot access. In that case, site the owner must take care about the improvement the structure of the site's URLs. Creating descriptive categories and filenames for the documents on a website can not only help keeping a site better organized, but it could also lead to better crawling

of documents by search engines. Also, it can create easier, "friendlier" URLs for those that want to link to your content.

Website security experts highlight the following statements which are very important for analysis,

1. Search spiders will not attempt to "submit" forms and thus, any content or links that might be accessible via a form are invisible to the engines.
2. Spiders will not attempt to perform searches to find content, and thus, it is estimated that millions of pages are hidden behind completely inaccessible internal search box walls.
3. It is recommended to use standard HTML links instead of Javascript based links on any page where search engine referred traffic is important.
4. Any links embedded inside Flash, Java applets, and other plug-ins are usually inaccessible to search engines.
5. The Meta Robots tag and the robots.txt file both allow a site owner to restrict spider access to the page.
6. The search engines all have a rough crawl limit of 150 links per page before they may stop spidering additional pages linked to from the original page.
7. Important pages may have upwards of 200 or even 250 links followed.
8. Sometimes nofollow expression is useful for security purposes: nofollow this link.

URLs with words that are relevant to your site's content and structure are friendlier for visitors navigating your site. Visitors remember them better and might be more willing to link to them. So URL Link names <http://www.imdb.com/title/tt0468569/?a=1> is less informative for users and search engines than <http://www.dmoz.org/Games/Video_Games/History/>.

Length of links, according Matt Cutt [5], leading google SEO guru, must be between 3 and 5 words, avoiding repetition. It is not clear about using synonyms in URL, and plural-singular words. Nobody can claim, that link <http://www.domain.com/horoscope-zodiac-horoscopes> is worse than <http://www.domain.com/zodiac-horoscopes>.

Most probably it depends on the site category and SEO situation, where synonyms are treated as equal words in exact topic. In many cases synonyms and plural-singular words can be treated as different words.

Also it is important to create a simple directory structure, to use a directory structure that organizes a content well and makes it easy for visitors to know where they are on your site. Navigation must be easy as well as site structure as presented in Fig. 4.

The navigation of a website is important to help visitors to find quickly the content they want. It can also help search engines understand what content the webmaster thinks is important.
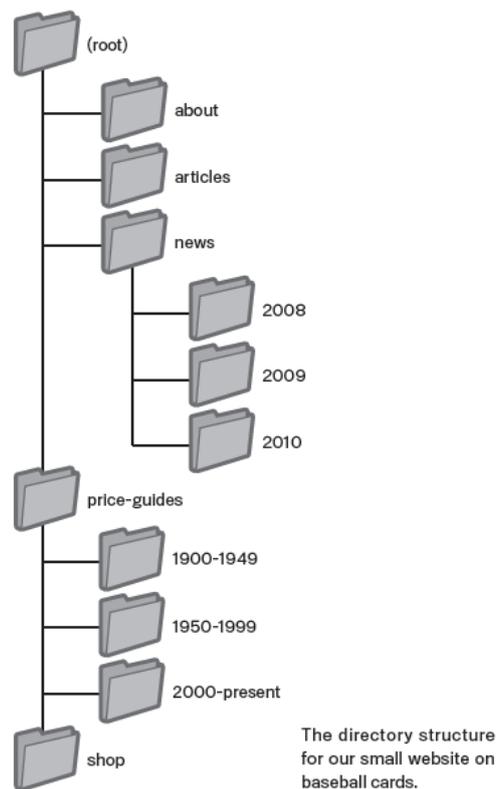

Fig. 4. Web site tree.

2.3. Domains

The importance of keywords in domain name plays the same role as in previous periods.

1. The benefit of a keyword-rich domain is two-fold. The domain name itself is a ranking factor that the engines consider when calculating ranking order.
2. Having relevant keywords in a domain name is beneficial because the domain name is the text that other Internet users will use as anchor text when linking. Since keywords in anchor text are an important ranking factor, having these keywords in a domain name can have a positive impact on ranking.
3. If your domain name is two words (like <www.examplesite.com>), you may want to separate the words with a hyphen for readability: <www.example-site.com>. So to say, the use of hyphens also correlates highly with spammy behaviour—and more than one hyphen should not be used in a domain name. For this reason, it is generally better to stick to the domain names containing only one or two words.

According to Matt Cutt [5], a domain name can have important keywords but it is not a strict rule. A lot of business web sites prosper without keywords. However, for a small site it can be one out of the main "jump to the top" factors.

2.4. Title Tag

Importance of <Title> tag (slightly decreased), but better to keep rules working on site. Some sites have no key keywords in the title but they are found on the first page. It means it has very important content:

1. Title tags longer than 70 characters may be truncated in the results or search engines may choose to display different text from the document in place of the title tag.
2. Optimal Format must be constructed using the following routine:

   Primary Keyword - Secondary Keyword | Brand Name
   Brand Name | Primary Keyword and Secondary Keyword
3. The main keywords must appear in the beginning of <title> tag. This is not a strict rule but this rule gives several advantages, for example, it is easy to test the simple entering search keyword on Google service.

2.5. Anchor text

Web site owners should care about links' text so called *anchor* text. It becomes especially important in the case of outbound links.

1. Anchor Text is the visible, clickable text in a hyperlink. In modern browsers it is often blue and underlined.
2. Link relevancy is determined by both, the content of the source page and the content of the anchor text. It is a natural phenomenon which occurs when people link out to other content on the web.
3. With the Penguin update, Google began to look more closely at keywords in anchor text.
4. If too many of a site's inbound links contain the exact same anchor text, it appears to be suspicious and may be a sign that the links were not acquired naturally.

2.6. Meta Description

Meta descriptions are commonly used on search engine result pages (SERPs) to display preview snippets for a given page.
```
<meta name="description" content="This is
an example of a meta description. ">
```
Optimal Length for Search Engines contains roughly 155 ascii characters. Meta description tags, while not important to search engine rankings, are extremely important in gaining user click-through from SERPs. These short paragraphs are a webmaster's opportunity to advertise content to searchers and let them know exactly whether the given page contains the information they are looking for. Meta tags are not a Google Ranking Factor. Google announced in September of 2009 that meta description and meta keywords do not affect the ranking algorithms of Google web search.

Fig. 5 represents well formulated structure of webpage. Some rules of usage of metadates metadatas are formulated below.

1. Avoid Duplicate Meta Description Tags and TITLE tags.
2. Sometimes it is good Not Write Meta Descriptions. Use the general rule that if the page is targeting between one and three heavily searched terms or phrases, go with a meta description that hits those users performing the search.

2.7. What is duplicated content?

The main purpose for a web site programmer could be formulated as follow: to create excelent and unique user friendly data content. New (and renewed) content will not only keep your existing visitor base coming back but will also bring in new visitors. Rehashing (or even copying) existing content that will bring little extra value to the users having duplicate or near-duplicate versions of your content across your site should be avoided.

When there are multiple pieces of identical content on the Internet, it is difficult for search engines to decide which version is more relevant to a given search query.

To provide the best search experience, search engines will rarely show multiple duplicate pieces of content and thus, are forced to choose which version is most likely to be the original, or the best.

Fig. 5. Well done structure of web page.

Table 1. Solution of task.

```
301 Redirect
Element <link href="http://www.example.com/canonical-version-of-page/" rel="canonical" />
Do not allow to index <head> <meta name="robots" content="noindex, follow" /> </head>
```

When duplicate content is present, the site owners suffer rankings and traffic losses, and search engines provide less relevant results.

Category URL systems create multiple versions of the same page, as presented below: <http://www.domain.com/quotes?type=funny> and <http://www.domain.com/funny-quotes-sayings-quotations>.

Print only URLs create multiple versions of the same content. Examples: <http://www.domain.com/funny-quotes-sayings-quotations> and <http://www.domain.com/print/funny-quotes-sayings-quotations>.

Session ID URLs multiple versions of the same content. Examples: <http://www.domain.com/print/funny-quotes-sayings-quotations> and <http://www.domain.com/print/funny-quotes-sayings-quotations?SESSID=1239834572>.

Other variables create multiple versions of the same content such as: <http://www.domain.com>, <http://www.domain.com?a=1>, <http://www.domain.com?a=2>, <http://www.domain.com?a=3>.

How to solve the task according to Ref. [6]? Solution is presented in Table 1.

2.8. What is robots.txt and site maps?

The robots exclusion protocol (REP), or robots.txt is a text file webmasters create to instruct robots (typically search engine robots) how to crawl and index pages on their website. Robots.txt needs to be placed in the top-level directory of a web server in order to be useful. Robots.txt file is useful because site owners can restrict we bots to access low level pages or pages which are necessary for site but works like "trash" in common web site content. For example: rules, disclaimer, and license. Table 2 represents the example of robots.txt file.

Table 2. Example of robots.txt file

```
http:/www.example.com/robots.txt

# Block all web crawlers from all content
User-agent: * Disallow: /

# Block a specific web crawler from a specific folder:
User-agent: Googlebot
Disallow: /no-google/

# Block a specific web crawler from a specific web page:
User-agent: Googlebot
Disallow: /no-google/blocked-page.html
```

This example contains restrictions, oriented to different search engines.

Sitemaps help to know importance of your pages. A site map (or sitemap) is a list of pages of a web site accessible to crawlers or users. It can be either a document in any form used as a planning tool for Web design, or a Web page that lists the pages on a Web site, typically organized in hierarchical fashion.

3. Tools

Here the question arises on how to know situation about the site and preferred, possible solution. In what state the site is now and what goals can be suggested? The answer is to use Google tools, which will help to understand exactly what is going on with the web site.

Google webmaster tools [3] is a web service by Google for webmasters. It allows webmasters to check indexing status and optimize visibility of their websites.

It has tools that let the webmaster:
a)   heck and set the crawl rate;
b)   list internal and external pages that link to the site;
c)   see what keyword searches on Google led to the site being listed in the SERPs, and the click through rates of such listings;
d)   wiew statistics about how Google indexes the site;
e)   submit and check a sitemap;
f)   generate and check a robots.txt file.

Google analytics [7] is the enterprise-class web analytics solution that gives you rich insights into your website traffic and marketing effectiveness. It let you see and analyze your traffic data, queries, visited pages.

Google Adwords Keyword Planner [8] is like a workshop for building new Search Network campaigns or expanding existing ones. You can search for keyword and ad group ideas, see how a list of keywords might perform, and even create a new keyword list by multiplying several lists of keywords together. These tools are free.

Conclusion

We can clearly understand that it is not simple to reach good position on Google. Many efforts must be put to reach first page or even first position on Google relying on exact keywords. Constant work must be done analyzing changes in SEO because Google-search and indexing algorithm have been changing and is becoming more and more sophisticated.

References

1.   Search Engine Optimization Starter Guide.   – <http://static.googleusercontent.com/media/www.google.com/en//webmasters/docs/search-engine-optimization-starter-guide.pdf>.

2.   Learn Seo – <http://moz.com/learn/seo>.
3.   Google Webmaster Tools – <www.google.com/webmasters/tools>.
4.   <https://www.google.com/webmasters/tools/disavow-links-main>.
5.   Google Webmasters – <http://www.youtube.com/user/GoogleWebmasterHelp>.
6.   Google Product Forums – <http://productforums.google.com/forum/#!forum/webmasters>.
7.   Google analytics – <http://www.google.com/analytics/ >
8.   Google Adwords Keyword Planner – <https://adwords.google.com/ko/KeywordPlanner/Home>.