INTERNATIONAL JOURNAL OF RESEARCH IN COMPUTER APPLICATIONS AND ROBOTICS

FACTORS AFFECTS SEO

[1]M. A. Abidou, [2]M. El-Kady, [3]A. S. Hussien
[1]Faculty of Engineering, Eng. Mathematics & Physics Department, AinShams University
[2]Faculty of Science, Mathematics Department, Helwan University. Egypt
[3]Faculty of Science, Mathematics Department, Helwan University. Egypt

[1]maged.abidou@eng.asu.edu.eg, [2]m_el_kady@hotmail.com, [3]ahmed.hussien.mail@gmail.com,

Abstract

Ranking is an important component of any information retrieval system. As the rapid growth of web, a simple key word search could correspond to hundreds of thousands of web pages. A human can usually verify that the twenty or a little more URL returned by the search engines. In this paper, we identify the most popular techniques used to rank a web page highly in Google.

Keywords: search engine; search engine optimization; web, SEO, factors.

1. Introduction

With the continued growth of the Internet and the amount of sites available, it has become increasingly difficult for sites seeking to gain public visibility. According to a recent study, there are about 3 million new sites on the Internet each month. [1] As a result of this continued growth, it has become increasingly difficult for websites to be visible to all the other competitors sites.

Another study found that "over 80% of first visits to a web site from search engine. Of those visits more than 76% use Google search in the world ". [2, 6] These studies show how to achieve top rankings in search engine results is essential to the continued visibility of a website. Therefore, get top positions in search engine results is essential for the constant flow of users to the websites, and this is where the value of SEO comes in.

2. Literature Review

Study conducted by Zhu (2011) used a reverse engineering method to "study and analyse the key factors that influence the search process on the Web." With this methodology, researchers identified five main factors for SEO. Researchers developed a system that crawl all site factors (e.g., HTML structure, the length of URL, etc...) for

200000 web pages with 10000 key search words as a whole sample. Moreover, the keywords in the set of samples were divided into the following three segments according to their Google search volume in the last three months: Hot "high search volume", Middle "medium search volume", and Cold "low search volume". The purpose of this classification was to discover the "different SEO factors in different segments" [2] of the keywords. Table 1 lists the five.

Table 1: The Top 5 SEO factors that used in Zhu'sstudy [2].

Rank	Hot	Middle	Cold
1	URL Length	URL Length	URL Length
2	Keyword appear in URL domain	Keyword appear in URL domain	Keyword appear in URL domain
3	Keyword density in H1	Keyword appear in URL path	Keyword density in title
4	Keyword density in title	Keyword density in title	Keyword density in title
5	URL layers	URL layers	Keyword density in H2

Another study by Su (2010) focused on the Google ranking algorithm, researchers systematically tried to "validate assumptions others have done about this Top ranking algorithm" and identify what page factors or any other criteria, had the greatest influence on their ranking algorithm. They designed and developed a ranking system to determine the most important factors that Google uses to rank pages.

Using a reversed engineering approach, paper showed how his classification system "can be used to reveal the relative importance of ranking factors according to ranking Google" [8]. Therefore, the document provides guidelines for SEO engineers what are the most important factors for the optimization of web pages for the purpose of achieve higher rankings.

Study conducted by Fuxue (2011) is based on the use of Chinese search engine Baidu, Google's PageRank algorithm is not considered. They picked up data from 116 web sites, the researchers tried to "analysing the impact of SEO techniques" and determine "what technique is most effective strategy" [9]. The authors determined a candidate metric, Interest page, and then investigated the effect of various SEO techniques in this metric. Interest page, you can see that if a large number of page views, an increase of time spent in the website and a low bounce rate, there will be a high interest page. The results of their study indicate that the page size, Customizing 404 pages and global links are important factors in the effectiveness of SEO.

3. How Google Works

When it comes to SEO Google is the search engine that matters most. Google has the lion's share of search traffic. Bing and Yahoo!, are increasing in importance, but still only control a relatively small percentage of search traffic.

There are three levels to the Google technology [7]:

- The "bot" or "spider" which crawls the web and indexes web pages
- The index which contains all of the information on a page
- The search query which end users enter into a search engine to find content
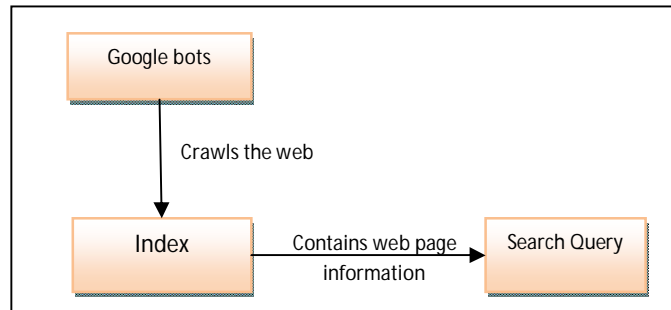
Figure 1: How Google works

Google indexes pages on a frequent interval. One way to ensure that the new content is indexed is to register a XML site map with instructions on how often to index a site and content to index.

Google returns web pages in a search in the order of their ranking of the page for that search term. Google uses a proprietary algorithm to calculate page rank. In generally Google have four factors to determine page rank:

1) The popularity of a web page roughly measured by how many sites link to it.
2) The authority of a web pages measured by:
   a. The age of the domain (an older domain is considered to have more authority)
   b. The top level domain (TLD). .Edu and .Gov websites have the most authority
   c. The popularly and authority of the external pages that that link to the page.
3) The relevance of a web page based on its contents and meta data or information about the information, as well as the relevance of the external or "inbound" links pointing to the page as expressed in the anchor text of those links.
4) The location of a web page based on where it lives in the overall information architecture of the site and how well it can be indexed by a search engine spider or bot.

4. SEO Approaches

In an SEO program there are two primary types of optimization: On-Page and Off-Page.
SEO FACTORS:

   I.   URL
A search engine friendly URL should be concise, and descriptive; e.g.
https://www.watsonhall.com/resources/downloads/top10-website-security-issues.pdf the use of hyphens between the terms is a best practice.
Many SEO experts think that hyphenated content ranks better than underscores. Connoted URLS without any space between terms will not rank as well.

Figure 2: Highly optimized URL.

In Figure 2 you can see a better structured and highly optimized URL. Users who come across similar URLs such as this will have a better idea of what the web page is about. This is helpful to both users and search engines, because self-explanatory and self-documenting URLs provide more useful information about the page and its content.

It is critical to avoid any special characters in a URL. A search bot will stop indexing when it hits a special character, automatically assuming there is not any valuable information following the character. For instance in the URL www.company.com/Admissions/?campus_tours=1
Google would not see any information after the "?" character.

II.   Robots meta tags

A robots meta tag is a line of HTML code that is included in a Web page to instruct search engine crawlers how they should process that page. Specifically, the tag tells crawlers if they are allowed to index the page, follow its links, and/or archive its contents. If you don't want to restrict the crawlers, you shouldn't include a robots meta tag.

III.   Use of robots.txt

The purpose of robots.txt is to signal the search engines whether to crawl your site and what locations (i.e. pages or directories) should not be accessed. Search engines send their crawlers every so often to crawl and index the content of web sites. And there may be times where when you want to keep certain directories or folders from getting crawled (for security reasons). In order to setup the permissions to certain directories or pages, you must configure the robots.txt file using the settings as seen in Figure 3.


```
www.google.com/robots.txt

User-agent: *
Disallow: /search
Disallow: /sdch
Disallow: /groups
Disallow: /images
Disallow: /catalogs
Allow: /catalogs/about
Allow: /catalogs/p?
Disallow: /catalogues
Disallow: /news
Allow: /news/directory
```

Figure 3: Sample configuration of robots.txt file and its location

Notice how the property User-agent is set to the wildcard * symbol; this configuration instructs all spiders (i.e. googlebot, msnbot, yahoo-slurp, etc…) to crawl and index all pages in the site. Furthermore, notice that the robots.txt file must be located in the root directory of the web server.

```
User-agent: *
Disallow:
```

IV.   XML Sitemap

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

V.   Page avoids using Flash

Flash creates pretty websites, but in most cases, those websites are virtually invisible to search engines.

VI.   Page contains substantive content

    A strategic way to structure the content of the website is to place the most important text on the top and should contain the target keyword optimization. Pages with substantive content are valuable to users and search engines because they provide sufficient information to answer search queries.

VII.   Page contains readable content

    If a page is easily readable, it will appeal to a large and diverse audience

VIII.   Page minimizes misspellings

If a page contains numerous misspellings, it loses credibility in the eyes of its audience.

IX.   Page's title is the appropriate length

Defines the title of the page and communicate what the page is about search engines, so that the target keyword should be inserted into this label.
The page title is the most critical piece of information in determining how well a web page will rank.

    A title should be 65 characters or less in length.

X.   Page's title is relevant for the page's content

It is important to have variation in your titles between pages. Each title should be unique. Google recommends choosing "a title that effectively communicates the topic of the page's content" [3] so that its search ranking algorithm can better categorize it and know what the page content is about.

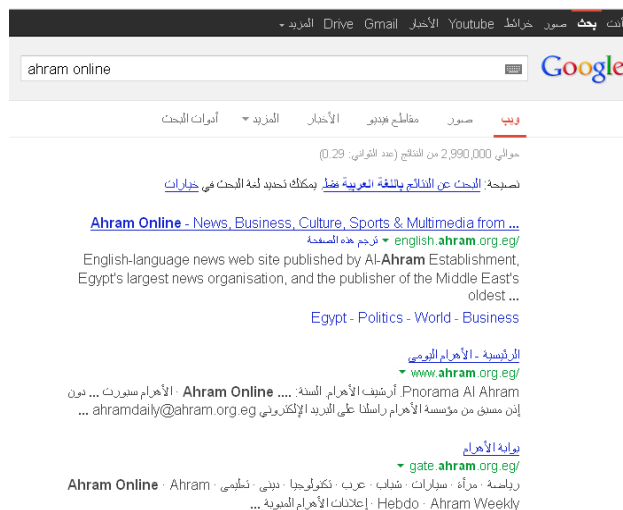Figure 4: The top three organic results for the keyword ahram online.

Figure 4 shows the top three organic results for the keyword ahram online. In this example, you can see that these top search results pages have been optimized for ahram online because words from my search query appear in bold within the hyperlink title. Having the keywords in the title tag, which appears as the hyperlink in Google's search

results, informs the searcher that if he or she clicks on the link, the resulting destination page will be relevant to the search query.

XI.   Page's meta description is the appropriate length

Search engines display a page's meta description in their results. If the description is too long, the additional characters will be cut off in the search results.

XII.   Page's meta description is relevant for the page's content

A Meta description describes the contents of a page and plays a role in page rank. Moreover, when Google lists a page in its results it often uses the Meta description as the abstract. Thus is important to write good Meta descriptions that include keywords, but more importantly provide a friendly description of the page. Google explains that "a page's description meta tag gives Google and other search engines a summary of what the page is about" [3].

You can see the use of the description meta tag in Figure 5 when a search for learn web design is performed. The words from the keyword are bolded or enhanced to give it its attention.
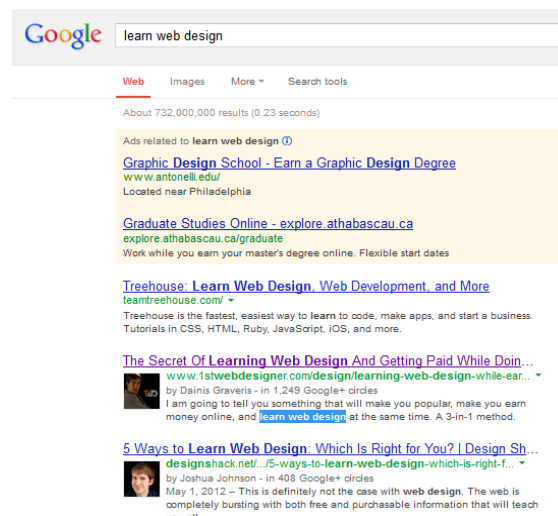

Figure 5: Use of the description meta tag in the search results

In Figure 6, you can see the optimization of the description meta tag as seen in the HTML source code of the page, notice how the target keyword (learn web design) is included in the description text.

Figure 6: Optimization of the description meta tag for learn web design.

XIII.   Page avoids using meta keywords

Meta Keywords are not used by search engines in their ranking algorithms any more. The one exception to this rule has been Yahoo!, which still indexes keywords, but they have very little impact on the actual ranking of content, rather are used as synonyms. However, with Microsoft Bing taking over Yahoo! Search it is unclear if keywords will be completely deprecated like they are on MSN. Many SEO experts recommend that site operators still use keywords.

XIV.   Page uses an H1 heading

A heading is the most prominent piece of content on a page. It carries weight in indexing. Heading tags are used to emphasize important text and structure on the web page. They also inform the search engines what your page is about; the <h1> tag defines the most important heading and <h6> defines the least important heading, so it's important to include your target keyword into the headings, most importantly within the <h1> tag. This can be effective because it helps establish information hierarchy within the page being optimized and it helps visitors understand the web page content and how it's organized.

Figure 7: H1heading optimization for learn html5 and how is viewed in the browser.

When optimizing the <h1> heading, it's recommended to use only one <h1> in the web page and include your target keyword once; using more than one <h1> heading and repeating your keyword multiple times can raise a red flag for a black-hat technique known as keyword stuffing, which can negatively affect your search rankings.

Figure 7 shows part of the HTML source code where the <h1> heading was optimized for my target keyword (learn html). Note the HTML source code and how the page is viewed in the browser.

XV.   Page's H1 heading is relevant for the page's content

These tags are used to highlight important text and structure of the website. Also inform the search engines what your page is about how it is organized. The tag defines the title <h1> more important and less important <h6> defines the starting.

XVI.   Page's images use alt text

The alt attribute from the <img> HTML tag specifies the alternative text of what the image is about in case the image doesn't load or cannot be displayed correctly. This short text description is helpful to the search engines and to people with disabilities so they can understand what's contained on the image.

In regards to SEO, Google takes the keywords from the alt attribute into account in their ranking algorithm. In addition, if an image is used as a link, "the alt text for that image will be treated similarly to the anchor text of a text link" [4]. So when optimizing your images, it's important to include your target keywords into the image alt description text.

Figure 14 shows the correct image optimization of the alt attribute; you can see how it's being optimized for the keyword: "html5". Also, it's recommended that you name your images with a descriptive name. And if possible, include your target keyword in the naming of the image as seen in Figure 8.

Figure 8: Image optimization example for the key html5

XVII.   Page's images are relevant for the page's content

Pages with relevant images are more engaging to their audience, and they convey visually consistent messages.

XVIII.   Page's URL uses hyphens instead of underscores

Search engines consistently treat hyphens as word separators, but not all of them use underscores for that purpose.

XIX.   Page's URL avoids using parameters

URL parameters introduce potential duplicate content issues, and they make the page less likely to be crawled by search engines.

XX.   Page's URL is relevant for the page's content

A relevant URL serves as an effective advertising tool, regardless of where it appears (e.g., search results, social media, etc.).

5. Experimental Results

In this study we analyzed, implemented and tested SEO factors. Moreover, their effectiveness in the ranking algorithm of Google search engines were tracked and measured. The following list provides a description of the SEO process applied in this research:

- Selecting a keyword "melgram"
- Keyword research: the effective keyword research is at the base of SEO. Keywords may be considered in the link that puts users to websites. This is why keyword research can be considered the most important part of the SEO process.
- Data collection: Setting up a monitoring tool for gathering website data. It is important to collect detailed statistics information about visitors to this site has been optimized, as it can provide useful data, such as how users come to your site, what keywords are responsible for sending the most visitors, what countries are sending the most visitors, etc… We selected first 20 results from Google search engine.

Table2-a: First 20 web pages analysed by SEO Tools

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12
Site 1	5	5	0	0	5	0	0	3	5	4	5	5
Site 2	5	5	5	0	5	1	0	4	5	5	0	1
Site 3	5	5	5	5	5	0	0	3	5	5	0	0
Site 4	5	5	5	0	5	0	5	4	5	2	0	1
Site 5	5	5	5	0	5	0	0	2	5	5	0	0

Site 6	5	5	5	0	5	4	0	5	0	2	0	3
Site 7	5	5	5	5	5	0	3	3	5	5	5	4
Site 8	0	0	5	0	5	0	0	0	5	5	5	5
Site 9	5	5	5	0	5	0	0	5	5	4	5	5
Site 10	0	5	5	5	5	5	0	0	5	3	5	2
Site 11	0	5	0	0	5	4	0	4	5	2	0	0
Site 12	5	5	5	0	5	5	5	2	5	4	5	5
Site 13	5	5	5	5	5	2	0	4	5	5	0	0
Site 14	0	3	5	0	5	4	0	5	5	4	5	3
Site 15	0	5	0	0	5	0	5	2	3	5	5	5
Site 16	5	5	5	0	5	3	0	5	5	5	0	5
Site 17	0	5	5	5	5	5	0	5	5	1	5	1
Site 18	5	5	5	5	0	5	0	3	5	2	4	0
Site 19	5	5	5	0	5	5	0	3	5	3	5	1
Site 20	5	5	5	0	5	2	0	5	5	3	4	0

Table2-b: First 20 web pages analysed by SEO Tools

	F12	F13	F14	F15	F16	F17	F18	F19	F20	Score
Site 1	5	1	0	0	5	0	5	5	5	58
Site 2	1	2	5	3	3	3	5	5	5	67
Site 3	0	5	5	5	1	4	5	0	5	68
Site 4	1	5	0	0	1	1	5	5	5	59
Site 5	0	5	0	0	0	0	5	5	5	52
Site 6	3	5	5	3	2	3	0	5	5	62
Site 7	4	5	5	5	1	5	5	5	5	86
Site 8	5	2	5	5	1	2	5	0	5	55
Site 9	5	5	0	0	5	0	0	0	5	59
Site 10	2	5	0	0	5	1	5	5	5	66
Site 11	0	5	0	0	5	0	5	5	5	50
Site 12	5	1	5	5	1	2	5	5	5	80
Site 13	0	5	5	5	1	1	5	0	5	68

Site 14	3	2	5	5	5	2	5	5	5	73
Site 15	5	5	5	5	4	1	5	5	5	65
Site 16	5	1	5	5	3	3	5	5	5	73
Site 17	1	2	5	1	4	0	5	0	5	64
Site 18	0	2	5	3	2	1	5	5	5	67
Site 19	1	1	5	5	3	1	5	5	5	75
Site 20	0	1	5	5	3	1	5	5	5	72

Factors weight was calculated by applying the following equation

$$f(W_n) = \sum_{f=1}^{n} (f_n * W_n)$$

(1)

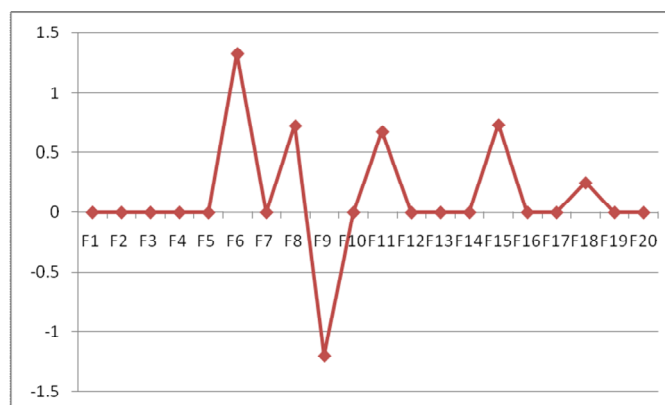In Equation (1), f (W) is a factor that weights the importance of factor (n).

Figure 9: Calculation of factors 'weight

In figure 9, the results shown that (F6) Page contains substantive content, is the most importance factor. Also (F15) Page's H1 heading is relevant for the page's content, (F8) Page minimizes misspellings, (F11) Page's Meta description is the appropriate length, and (F18) Page's URL uses hyphens instead of underscores have more value than other factors. While (F9) Page's title is the appropriate length, is less important than other factors.

6.   Conclusion

Search Engine Optimization is a keystone of any good web marketing strategy. The good SEO skills are also good web management skills, as they give rise to the application based on web standards, information organization, and well written, the subject content.

Finally, sites must do what is necessary to get noticed the competition in order to stay in business, which means start paying attention to key processes such as SEO in order to increase your website visibility.

SEO methods will keep growing in relevance and research in this area is bound to grow as the Internet continues to grow. As mentioned, Google has become an essential link in the process through which clients and companies to find each other. And in many ways, SEO can help close that gap.

7. Acknowledgment

8. References

[1]   A. Grzywaczewski. "E-Marketing Strategy for Businesses" e-Business Engineering (ICEBE) IEEE 7th International Conference on, E-ISBN 978-0-7695-4227-0, 2010, pp. 428 - 434.

[2]   C. Zhu. "Research and Analysis of Search Engine Optimization Factors Based on Reverse Engineering" Multimedia Information Networking and Security (MINES) Third International Conference ISBN 978-1-4577-1795-6, 2011, pp. 225 - 228.

[3]   Google, (2013), "Search Engine Optimization - Starter Guide", [Online]. Available: http://www.google.com/webmasters/docs/search-engine-optimization-starter-guide.pdf [May. 30, 2013]

[4]   D. Dover, E. Dafforn. "Search Engine Optimization (SEO) Secrets" Wiley Publishing, Inc., Indianapolis, Indiana. ISBN: 978-0-470-55418-0, 2012

[5]   SEOMOZ. (2012), "The beginner guide to CEO", [Online]. Available: http://static.seomoz.org/files/SEOmoz-The-Beginners-Guide-To-SEO-2012.pdf [May. 30, 2013]

[6]   K. Purcell, J. Brenner, and L. Rainie. (2012), "Search Engine Use 2012", [Online]. Available: http://www.pewinternet.org/Reports/2012/Search-Engine-Use-2012/Summary-of-findings.aspx [May. 30, 2013]

[7]   G. Lu. "Multimedia Database Management Systems" Artech House, Boston, London, ISBN 1-89006-342-7, 1999, pp. 96-99

[8]   S. Latha, M. Rajaram. "Semantic Based Multiple Web Search Engine " (IJCSE) International Journal on Computer Science and Engineering, Vol. 02, No. 05, 2010, pp. 1722-1728

[9]   K. Purcell, J. Brenner, and L.Rainie (2012), "Search Engine Use 2012", [Online]. Available: http://www.pewinternet.org/Reports/2012/Search-Engine-Use-2012/Summary-of-findings.aspx, [May. 30, 2013]