UDC 001.81:004.774
Original Scientific Paper

Measuring the Utilization of On-Page Search Engine Optimization in Selected Domain

Goran Matošević *gmatosev@unipu.hr*
Faculty of Economics and Tourism "Dr. Mijo Mirković"
University of Pula, Pula, Croatia

Abstract

Search engine optimization (SEO) techniques involve „on-page" and „off-page" actions taken by web developers and SEO specialists with aim to increase the ranking of web pages in search engine results pages (SERP) by following recommendations from major search engine companies. In this paper we explore the possibility of creating a metric for evaluating on-page SEO of a website. A novel „k-rank" metric is proposed which takes into account not only the presence of certain tags in HTML of a page, but how those tags are used with selected keywords in selected domain. The „k-rank" is tested in domain of education by inspecting 20 university websites and comparing them with expert scores. The overview of results showed that „k-rank" can be used as a metric for on-page SEO.
Keywords: Search engine optimization, on-page SEO, website ranking, k-rank, website keywords

1. Introduction

Today search engines are one of most used tools to search for information and websites of interest. When searching, users are using keywords – terms of one or more words that describe the searched topic. Search engines are returning a list of websites that according to their algorithms match entered keywords [5]. This returned ordered list is called „search engine result pages" (SERP) and contains ranked list of websites. SERP is also called „organic search results" or „natural search results" – meaning they are constructed based on automatic process conducted by search algorithm. Besides organic list, search engines return advertisement list, or paid listings, containing list of websites which paid to be on the list for particular keyword. Paid listing is also ranked based on keyword bids and other ad quality factors (keywords in ad text, ad relevance, landing page quality etc.). This kind of advertising on search engines is based on PPC model (pay per click). It is clear that being on top of organic or paid SERP will lead to increase of website visibility and gain more visits [17].

While coming onto paid SERP is relatively easy, gaining top ranking in organic SERP requires more time and effort. This process is called „search engine optimization" (SEO) and is also referred as „free" search engine marketing (you can't buy organic SERP rank, you have to earn it by building high quality pages). Both, PPC and SEO are part of search engine marketing, and in broader context, of Internet marketing.

This paper explores main „on-site" factors that webmasters[1] use to conform to search engines guidelines and proposes a new metric called „k-rank" that indicates to which extent a website utilizes on-page SEO. Many SEO tools today for checking on-page SEO of a website report only keyword density in each element of the page (title, meta description, headings etc.) making it hard to do an analysis of more websites and more similar keywords. The novelty of „k-rank" lies in using domain keywords and weights to produce unique score which can be used to rank and compare websites.

[1] A person responsible for maintaining a website, also called web developer, site author or website administrator.

2. SEO techniques

SEO is a process of following search engines guidelines [1,7] while building websites. It includes various „on-site“ and „off-site“ techniques that influence the SERP ranking [15]. The aim of SEO is to rank high in organic SERP for particular keyword(s).

2.1. On-page SEO

On-page SEO includes techniques that webmaster can use directly on their webpages. This involves using particular HTML tags and keywords on the right places. On-page SEO is in total control of webmaster, and depends only on webmasters skill and knowledge. There are many on-page factors that search engines take into account while crawling and ranking web pages (more then 200), but according to the guidelines of the main search engine companies [1,7,15,18], the most important ones, which should contain targeted keywords are:

- Title tag
- Meta description tag
- Heading tag (h1)
- URL
- ALT tag
- Page content

Webmaster can influence how their webpages are showed in SERP by using title and meta tags, and using keywords in URL.

The HTML “title” tag is displayed on SERP as a link, usually with larger font. It’s the first thing that search engine users see while browsing the SERP pages.

Below the “title” on SERP there is a short textual description of the page. This description is pulled out from “meta description” tag of HTML. If a page does not contain meta description tag the description is generated from main page text (“body” tag).

SERP is also displaying the URL of the landing page. Having the right keywords in both places positions the website higher in SERP. Also, search engines tend to bold keywords in SERP making it more visible and generating more CTR[2].

It is therefore clear that webmaster and SEO specialists must know how to choose the right keywords and optimize the content and meta information of the website accordingly. Keyword research can be done by using several tools that exist on the market. Google is offering “Keyword planner”[3] as part of their advertisement program AdWords.[4] Other companies that offer SEO services also offer various tools for keyword research. Keyword research is all about choosing the right keywords, with low competition, but high search volume. Keyword density in web page’s content and variety (synonyms) is one of important on-page SEO factors.

Using keywords in page headings (tags H1, H2) and in page content is good practice. Images on page should contain ALT tag describing the image with selected keywords.

These are most important on-page SEO factors. Most SEO tools on the web are using these factors when checking for website optimization. There are a lot of more on-site factors that affect ranking, but their impact is lower.

Beside SEO positive factors, there are also negative ones, which can cause penalties, rank drops and even dropping out of search engine index. Overusing of keywords, known as “keyword stuffing” (too high keyword density) is one of those signals. Using hidden text or a page with too many ads are also negative factors. Search engines are trying to detect those factors to identify low quality and spam pages.

[2] The click through rate (CTR) is the number of times a click is made on the advertisement divided by the total impressions.
[3] https://adwords.google.com/KeywordPlanner
[4] AdWords is an online advertising service for search marketing by Google. Accessible at: http://adwords.google.com

2.2. Off-page SEO

While on-page SEO is all about what you can do with your website to rank higher in SERP, off-page SEO is what others says about your website. Until Google's PageRank algorithm on-page SEO was one of main ranking factors. PageRank brought revolution in search engine ranking algorithms by introducing link analysis as main ranking factor [10,11]. The main idea is to use links to websites as votes – website with more inbound links is considered more important. It's not only the number of links that counts, but also their origin – links from better websites (with high quality content) are counted as more important then links from smaller websites with low quality content. One of most important factor in links is anchor text – the text of the link. If the anchor text contains targeted keywords and the links comes from authority web sites, this is a strong signal of website quality and popularity.

How often is webpage's content shared on respected social networks is also a ranking factor. Many companies today have profiles on social networks trying to connect more closely with their potential customers. If real people "like" and share links, pictures and other website content these links count for ranking in SERP.

Other off-page ranking factors include domain history, location, author and others.

Search engine companies do not reveal the details behind their ranking algorithms. They only release general guidelines for webmasters and SEO specialists that should be followed if one wants to increase their position in SERP. The quality of webpage content is what should be in the spotlight and links will come naturally.

As there are negative signals in on-page SEO, so they exist in off-page SEO. Too many links in short time period is one of them. Links from spam websites is another negative factor.

3. Related work

In [2, 9, 12, 14, 17] authors explain what SEO is and how to use it. The search engine companies publish their own guides [1,7] on how to use SEO techniques to gain higher rankings in their search applications.

Similar researches investigating website rankings in SERP and factors that influence the position have been conducted in [4]. Authors investigate 50 highly optimized webpages from SEO competition and their ranking in Google to extract key factors and techniques influencing their rank. The paper gives insights into the off-page SEO techniques used by experts to rank highly in Google. The on-page factors was not taken into account.

Authors of [5] investigated the impact of PageRank on ranking in SERP, but as in previous research only links that affect PageRank were taken into account, and PageRank is not the only factor in ranking algorithm [13].

In [8] authors have analyzed selected Croatian faculty websites according to several on-page and off-page factors. However, their research does not take into consideration the usage of keywords in optimization which we believe is essential in reviewing the quality of SEO on inspected websites.

The importance of usage of keywords is explored in [16]. According to these authors it's important to use keywords on these places in HTML document: in page title, in meta description, in heading tags, in body of the page (keyword density), and in „alt" attributes of images. In [3] authors lists other factors as important ones for „on-page" SEO like PageRank, keywords in domain name, domain age etc.

In [19] authors are using reverse engineering techniques to extract important ranking factors. They crawled 200 thousand web pages and done content analysis based on top 20 positions of Google search result pages. As result they derived top five factors: URL length, keywords in URL, keywords in H1 tag, keywords in title and URL layers. These factors are also used in this paper.

4. K-rank methodology

The elements of k-rank are showed on Figure 1.

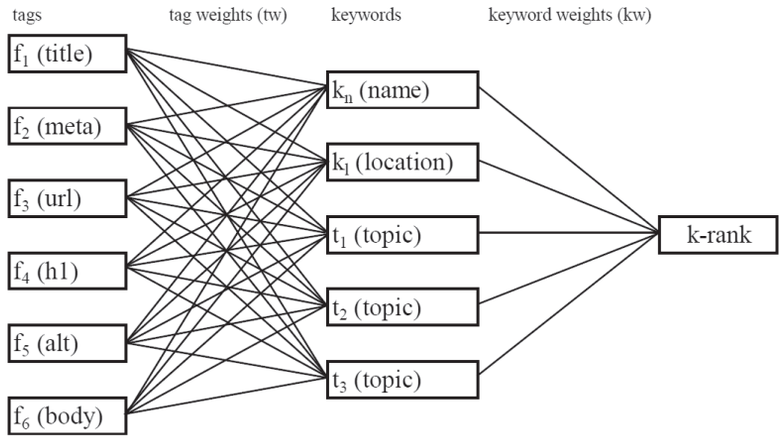


Figure 1. K-rank construction

We selected 6 most influenced on-page factors:

1) Title - The <title> tag is required in all HTML documents and it defines the title of the document. Its content is displayed in the browser toolbar, in favorites and in SERP. HTML example:

<title>University of Zagreb</title>

2) Meta - The <meta> tag provides metadata about the HTML document. Metadata is not displayed on the page, but is machine parsable and is used by search engines. The meta description tag is used for providing short description of the webpage content. Search engines display this short text on SERP. HTML example:

<meta name=”description” content=”Web pages of University of Zagreb - the oldest and biggest university in South-Eastern Europe.”>

3) URL - Usage of keywords in URLs, for example following URL contains keywords “university” and “research”:

www.unizg.hr/university/research.html

4) H1 - The <h1> to <h6> tags are used to define HTML headings. The <h1> defines the most important heading. HTML example:

<h1>University of Zagreb</h1>

5) Alt - The alt attribute is used in HTML IMG tag to specify alternative text (alt text) that is to be rendered when the image in IMG tag cannot be rendered. HTML example:

6) Body - The <body> tag defines the document's body. The <body> element contains all the contents of an HTML document, such as text, hyperlinks, images, tables, lists, etc. HTML example:

<body> Welcome to University of Zagreb </body>

From SEO aspect it’s important to have keywords in each above element. Factors (elements) are weighted based on their importance from 0-1 as shown in Table 1. They are set empirically based on popular search engines webmaster guidelines [1,7,15].

	f_1	f_2	f_3	f_4	f_5	f_6
parameter:	title	meta	url	h1	alt	body
weight:	1.0	0.2	0.3	0.9	0.2	0.5

Table 1. Tag weights

We evaluate each factor against 5 selected keywords in selected domain. The keywords are selected as follows: one keyword is the name (title) of the website inspected, one is the location of the company presented on website (city), and other 3 are most frequent topic

keywords in selected domain. Each keyword is also weighted according to importance for analysis (0-1).

The formula of k-rank in eq.1 is linear combination of tags (factors $f_1,..,f_6$), tag weights ($tw_1,..,tw_6$), keywords (k_n, k_l, t_1, t_2, t_3) and keyword weights (kw_n, kw_l, kw_{t1}, kw_{t2}, kw_{t3}) where:

$f_1,..,f_6$ – is number of times a keyword (k_n, k_l, t_1, t_2, t_3) appears in a tag f;

$tw_1,..tw_6$ – weights for each tag – the importance of tag containing the keyword with range between 0-1;

k_n – name keyword – the name of the website (company, organization or logical site title);

k_l – location keyword – the name of the city, region or country for which the site should appear in local search;

$t_1,...,t_3$ – topic domain keywords – keywords users use to find information in domain of the website analyzed;

kw - keyword weight – the importance of keyword in selected domain;

α - alfa parameter – bias or fixed parameter.

$$\text{k-rank} = \alpha + \sum_{x=kn,kl,t1,t2,t3} (\sum_{y=1}^{6} f_y tw_y) kw_x \quad (1)$$

where $f_y \in \{0,1\}$, depending on if the keyword is found (value 1) or not (value 0) in particular tag.

Changing the weights affects the k-rank. The weights should be set by SEO professionals according to their best knowledge and experience. Besides tag weights, alfa parameter must also be set. This parameter depends on scale that we want to use and is set empirically. In our example and k-rank evaluation we use the scale from 0.00 to 5.00, in which case we choose to set α parameter to 0.2. Tag weights, or importance of having keywords inside these tags, are set by SEO experts.

The weights must be set for keywords too. The idea is to use the same keywords and weights in entire research in selected domain. Only this way the calculated k-rank can be used as an indicator of on-page SEO quality. We even believe that when comparing k-rank of websites from different domains calculated with different keywords, as long the weights stays the same, can be a good metric for comparison.

5. Evaluation

For evaluation of proposed k-rank metric we randomly picked 20 websites from the list of university websites from UK[5]. Each website was manually inspected by 3 independent SEO experts and scored on the scale from 0.00 to 5.00, where 0.00 is the worst score and 5.00 is the best score regarding on-page SEO utilization. When scoring, SEO experts did not know the keywords we selected. We then computed the mean score for each website which we later compare with k-rank score. The results of experts scoring is presented in Table 2.

Website	Expert 1	Expert 2	Expert 3	Mean
1	3.90	3.50	2.50	3.30
2	2.70	3.50	2.50	2.90
3	3.00	3.75	3.50	3.42
4	3.50	3.60	3.70	3.60
5	3.80	4.00	3.00	3.60
6	4.00	3.50	4.00	3.83
7	3.90	3.00	4.00	3.63

[5] http://www.webometrics.info/en/Europe/United%20Kingdom

8	3.70	4.00	3.80	3.83
9	2.50	3.00	2.00	2.50
10	3.40	4.50	3.70	3.87
11	1.50	2.00	1.50	1.67
12	2.00	3.50	2.00	2.50
13	3.60	4.50	3.80	3.97
14	2.50	2.70	3.20	2.80
15	3.00	4.00	3.20	3.40
16	2.00	1.00	2.00	1.67
17	2.50	0.75	1.00	1.42
18	3.00	4.00	3.00	3.33
19	3.50	2.25	2.50	2.75
20	3.50	4.15	4.00	3.88

Table 2. Experts on-page SEO scores for selected 20 websites

For the purpose of this research and higher education domain we selected following keywords for computing k-rank:

kn=”university”
kl=[city of the university]
t1=”college”
t2=”study”
t3=”research”

These are the keywords we believe every university website should contain. The importance of these keywords are set with keyword weights as following:

kwn=1.0
kwl=0.8
kwt1=0.5
kwt2=0.1
kwt3=0.2

When we select keywords, we must comfort to certain rules and best practices. Keywords should be relevant to entire domain – in our example that’s higher education domain. Also, keywords must be used in search queries. To ensure that, we used Google Keyword planner and Google Trends tool[6]. Keyword planner is a tool by Google available in its Google Adwords advertisement program and is used to find keyword ideas in certain domain for advertisement on SERP. For each keyword the tool is giving monthly search volume estimate. By inspecting Google Keyword Planner in the category “Colleges, Universities & Post-Secondary Education” and in search territory “United Kingdom” we have results presented in Table 3.

Keyword	**Avg. monthly searches**
Ucas	1000000
Ise	165000
Cardiff portal	135000
Ucl	90500
Loughborough university	90500
Newcastle universtiy	74000
Cardiff universtiy	60500
Coventry universtiy	60500

[6] https://www.google.com/trends

Nottingham university	60500
Warwick university	60500

Table 3. Results of keyword research in Google Keyword Planner tool for category "Colleges. Universities & Post-Secondary Education" for UK

From Table 3 we can see how important is using the university name and location in SEO. That's why we have "name keyword" and "location keyword" in k-rank formula with high weights values. For setting the keyword weights for t_1, t_2 and t_3 Google Trends tool helped us. Inspecting selected keywords in Google Trends with location setting set to "United Kingdom" the "Interest over time" graph presented in Figure 2 helped us to determine keyword importance for each searched term.

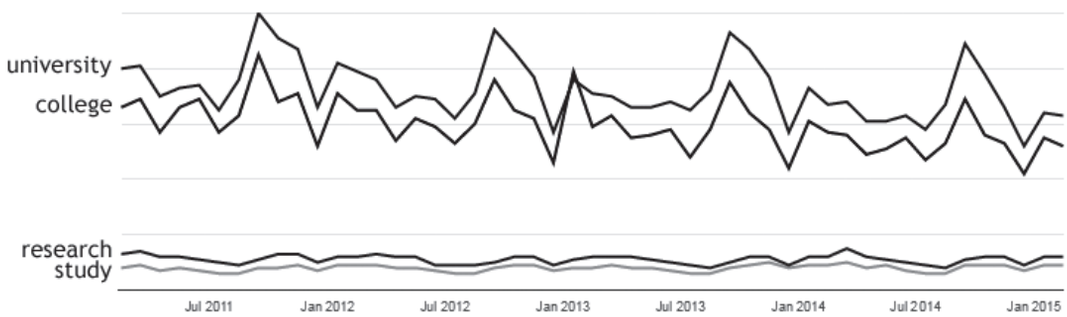
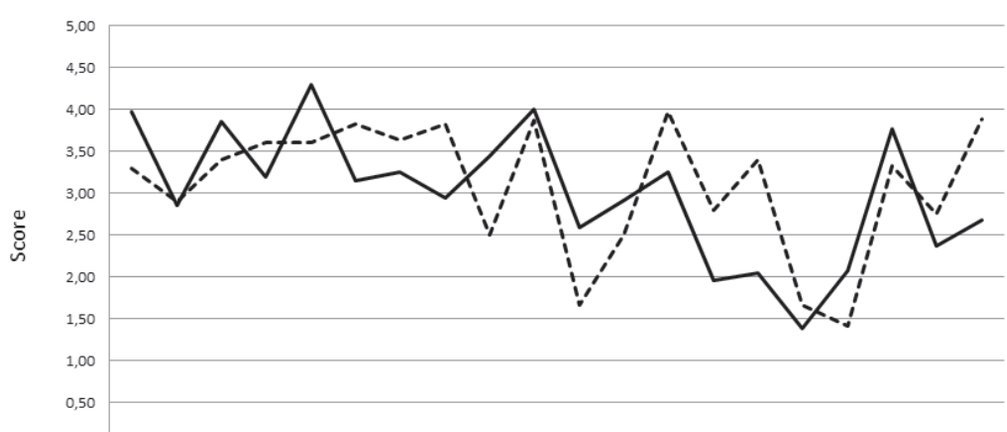


Figure 2. Interest over time for selected keywords in Google Trends tool for location „United Kingdom"

By using weights from Table 1 we can calculate maximum k-rank. In our case this is 7.44, which can be use to normalize k-rank. Since the experts scored the selected websites in scale from 0.00 to 5.00 we also have set the k-rank to be in this range. The comparison of calculated k-rank and mean expert scores are given in Figure 3.



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
k-rank	3,97	2,85	3,85	3,19	4,29	3,16	3,26	2,95	3,44	4,00	2,60	2,91	3,26	1,96	2,05	1,39	2,08	3,77	2,37	2,68
expert	3,30	2,90	3,40	3,60	3,60	3,83	3,63	3,83	2,50	3,87	1,67	2,50	3,97	2,80	3,40	1,67	1,42	3,33	2,75	3,88

Figure 3. Comparison of k-rank and mean expert score for selected 20 websites

Mean squared error between k-rank and mean expert score is 0.50, mean absolute error is 0.62 and the bias -0.09. We consider these statistics to be in acceptable range and conclude that k-rank does a good job in differencing websites by on-site SEO usage.

6. Conclusion

This research overviewed top on-page factors and proposed a k-rank methodology to quantify the on-page factors in combination with search keywords. The result showed that proposed k-rank can be a good metric for measuring the quality of on-page optimization. By adjusting the weights researcher can fine tune the k-rank to fit their own believes on importance of particular tags and keywords. Even this research proposes 6 tags and 5 keywords, further research can be done by using more tags and keywords potentially giving more detailed results. The k-rank could be improved by using other on-page factors like site speed, usage of keywords in links, content freshness, code vs. content ratio etc. Also, the k-rank formula (1) does not take into account multiple keyword occurrences (term frequency) in tags. Further research could try to include this feature, as it can be good SEO indicator especially in factor 6 (body of the page).

The evaluation of k-rank is limited by small sample and only 3 experts evaluators. Further research should include more websites from different domains (e.g. business, games, science, sports etc.).

References

[1] Bing webmaster guidelines. http://www.bing.com/webmaster/help/webmaster-guidelines-30fba23a, downloaded: May, 05th 2014.

[2] Chojar, B; Chauhan, A; Singh, A. V. Search engine optimization: A review. *International Journal of Managment, IT and Engineering*, 4(1): 493-503, 2012.

[3] Ellis-Chadwick, F; Mayer, R; Johnston, K. *Internet marketing: strategy, implementation and practice*. Pearson Education, 2009.

[4] Evans, M. P. Analysing Google rankings through search engine optimization data, *Internet research*, 17(1): 21-37, 2007.

[5] Fortunato, S; Boguna, M; Flammini, A; Menczer, F. How to make the top ten: Approximating PageRank from in-degree. *arXiv preprint*, cs/0511016, 2005.

[6] Gandour, A; Regolini, A. Web site search engine optimization: a case study of Fragfornet, *Library Hi Tech News*, pp. 6-13, 2011

[7] Google Search Engine Optimization Starter Guide. http://static.googleusercontent.com/media/www.google.com/en//webmasters/docs/search-engine-optimization-starter-guide.pdf, downloaded: May, 05th 2014.

[8] Gregurec, I; Grd, P. Search Engine Optimization (SEO): Website analysis of selected faculties in Croatia. In *Proceedings of Central European Conference on Information and Intelligent Systems*, pages 211-218, Varaždin, Croatia, 2012.

[9] Gupta, S; Aggarwal, A. Study of Search Engine Optimization. *International Journal of Research in Engineering & Applied Sciences*, 2(2):1529-1536, 2012.

[10] Is Google PageRank still important in Seach Engine Optimization? http://www.webseoanalytics.com/blog/is-googlepagerank-still-important-in-seach-engineoptimization/, downloaded: May, 05th 2014.

[11] Langville, A. N; Meyer, C. D. *Google's PageRank and beyond: The science of search engine rankings*. Princeton University Press, USA, 2011.

[12] Malaga, R. A; Worst practices in search engine optimization. *Communications of the ACM*, 51(12):147-150, 2008.

[13] Moran, M; Hunt, B. *Driving Search Traffic to your Company's Web Site*. Search Engine Marketing, Inc., IBM Press, USA, 2006.

[14] Patil Swati, P; Pawar, B. V; Patil Ajay, S. Search Engine Optimization: A Study. *Research Journal of Computer and Information Technology Sciences*, 1(1):10-13, 2013.

[15] Search Engine Land's Guide To SEO. http://searchengineland.com/guide/seo, downloaded: May, 05th 2014.

[16] Su, A. J; Hu, Y. C; Kuzmanovic, A; Koh, C. K. How to improve your Google ranking: Myths and reality. In *Proceedings of Web Intelligence and Intelligent Agent Technology (WI-IAT)*, on pages: 50-57, 2010.

[17] Yalçın, N; Köse, U. What is search engine optimization: SEO?. *Procedia-Social and Behavioral Sciences*, 9:487-493, 2010.

[18] Zhang, J; Dimitroff, A. The impact of webpage content characteristics on webpage visibility in search engine results (Part I). *Information Processing & Management*, 41(3):665-690, 2005.

[19] Zhu, C; Wu, G. Research and Analysis of Search Engine Optimization Factors Based on Reverse Engineeing. In *Proceedings of International Conference Multimedia Information Networking and Security (MINES)*, pages 225-228, 2011.