Towards a metric for on-page search engine optimization

Goran Matošević
Faculty of Economics and Tourism "Dr. Mijo Mirković"
University of Pula
Preradovićeva 1/1, 52100 Pula, Croatia
{gmatosev}@unipu.hr

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 ranking and comparing university websites from Croatia and UK. The overview of results showed that „k-rank" can be used for rank and comparison of how websites utilize 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 [4]. 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 [16].

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 number which can be used to rank and compare websites.

2 SEO techniques

SEO is a process of following search engines guidelines [5,10] while building websites. It includes various „on-site" and „off-site" techniques that influence the SERP ranking [13]. 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

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

pages (more then 200), but according to the guidelines of the main search engine companies [5,10,13,16], 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.

2.2 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 [8,14]. 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 [1 ,12, 7, 9, 15] authors explain what SEO is and how to use it. The search engine companies publish their own guides [5,10] 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 [2]. Authors investigate 50 highly

[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

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 [3] 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 [11].

In [6] 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.

4 K-rank methodology

The elements of k-rank are showed on Fig. 1. We selected 6 most influenced on-page factors:

- 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>

- 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.”>

- URL - Usage of keywords in URLs, for example following URL contains keywords “university” and “research”:
www.unizg.hr/university/research.html

- 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>

- 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:

- 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. 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).

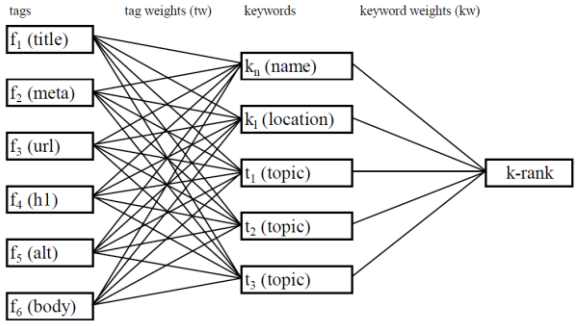

Figure 1. K-rank construction

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.

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

Changing the weights we can affect the k-rank. The weights should be set by SEO professionals according to their best knowledge and experience. Besides tag weights, tag thresholds must be set. A threshold is the upper limit, the number of times a keyword can appear in a tag that is not considered as

spam or "black hat SEO".[5] We set these parameters as in Table 1.

Table 1. Tag weights and tresholds

parameter:	f_1 title	f_2 meta	f_3 url	f_4 h1	f_5 alt	f_6 body
weight:	1.0	0.8	0.6	0.7	0.5	0.2
threshold:	2	4	2	3	4	15

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 Comparing university websites in Croatia and UK using k-rank

In our research of university websites in Croatia and UK we chose to set the weights as shown in Table 1.

For Croatian keywords we selected following:

k_n="sveučilište"[6]

k_l=[city of the university][7]

t_1="fakultet"[8]

t_2="studij"[9]

t_3="istraživanja"[10]

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

kw_n=0,8

kw_l=0,5

kw_{t1},..,kw_{t3}=1,0

For analysis of UK's university websites we used English equivalent keywords for k_n, t_1,..t_3:

k_n="university"

k_l=[city of the university]

t_1="college"

t_2="study"

t_3="research"

The same tag and keyword weights are used for both analyses. By using above weights and thresholds from Table 1. we can calculate maximum k-rank. In our case this is 58.1, which can be use to normalize k-rank. The results with normalized k-rank are summarized in Table 2 and Table 3. K-rank marked with (*) indicates that some of f parameters were above specified threshold. For that parameters the maximum threshold value vas used instead the real value.

Table 2. On-page SEO analysis of Croatian university websites according to k-rank methodology

Website	Keywords	f_1	f_2	f_3	f_4	f_5	f_6	k-rank
unizg.hr	sveučilište		1				5	0.039
	fakultet							
	studij							
	istraživanja						1	
	zagreb		1				2	
uniri.hr	sveučilište		1				3	0.115
	fakultet					4		
	studij							
	istraživanja						1	
	rijeka		2				13	
unios.hr	sveučilište	1					6	0.123*
	fakultet						11	
	studij							
	istraživanja							
	osijek	1					15	
unist.hr	sveučilište	1	1		1	1	7	0.137
	fakultet						2	
	studij							
	istraživanja					1	1	
	split	1	1		1	4	11	
unizd.hr	sveučilište	1						0.022
	fakultet							
	studij							
	istraživanja							
	zadar	1						
unidu.hr	sveučilište	1	1			1	12	0.166*
	fakultet							
	studij		1				15	
	istraživanja							
	dubrovnik	1	1				12	
unipu.hr	sveučilište				1		4	0.099
	fakultet						2	
	studij				1		10	
	istraživanja						4	
	pula				1		3	

Table 3. On-page SEO analysis of selected UK university websites according to k-rank methodology

Website	Keywords	f_1	f_2	f_3	f_4	f_5	f_6	k-rank
cam.ac.uk	university	1	4			1	10	0.279
	college						4	
	study		1				3	
	research		4				11	
	cambridge	1	3			1	13	
ox.ac.uk	university	1	2		1		5	0.187*
	college						15	
	study						2	
	research						5	
	oxford	1	1		1	1	15	
ucl.ac.uk	university	1	2		1		4	0.128
	college		1				1	
	study							
	research						5	
	london	1	2		1		5	
ed.ac.uk	university	1	2			1	5	0.153
	college							
	study						2	
	research		2			1	5	
	edinburgh	1	1			2	7	

[5] "Black hat SEO" attempts to improve rankings in ways that are disapproved of by the search engines, or involve deception.

[6] eng.: university

[7] The name of the city where the university is located

[8] eng.: college

[9] eng.: study

[10] eng.: research

nottingham .ac.uk	university	2			1		15	0.181
	college							
	study						7	
	research					2	9	
	nottingham	1		1	1		6	
gla.ac.uk	university	1	3		1	3	5	0.255
	college						1	
	study						5	
	research		3				10	
	glasgow	2	2		3	1	8	
southampto n.ac.uk	university	1	1				6	0.098
	college							
	study						2	
	research						5	
	southampto n	1	1	1			7	

By analyzing the above results we can see how specified parameters affect the k-rank. After manual inspection of selected websites we 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 and thresholds 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.

The research of Croatian and UK university websites is limited by small sample. Further research should include more websites and different domains (e.g. business, games, science, sports etc.).

The k-rank can be used not only to compare websites on-page SEO in certain domains, but also to detect spam websites by setting keyword count thresholds.

References

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

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

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

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

[5] Google Inc., 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.

[6] 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.

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

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

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

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

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

[12] 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.

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

[14] WebSEO analytics blog, 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.

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

[16] 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.