Using anchor text to improve web page title in process of search engine optimization

Goran Matošević

Juraj Dobrila University of Pula

Faculty of economics and tourism "Dr. Mijo Mirković"

P. Preradovića 1/1, 52100 Pula, Croatia

`{gmatosev}@unipu.hr`

Abstract. *Search engine optimization involves "on-page" and "off-page" techniques which webmasters use to improve ranking of web pages in popular search engines. One of main "on-page" ranking factor is page title which search engines display as link in their search results pages. Using keywords in page title is one of most important task of SEO. This paper proposes using top anchor text from backlink analysis to improve page titles of existing web pages. An experiment has been done on DMOZ websites and the results shows that anchor text can be used to improve page titles in process of SEO. This technique can be used to automate SEO process.*

Keywords. Search engine optimization, web page title, anchor text, backlink analysis

1 Introduction

For retrieval and ranking of web pages in search results popular search engines use various techniques in their algorithms. Search engines try to deliver most relevant results to user query and present relevant web pages in their SERP (Search engine results pages). Webmasters (a person responsible for maintaining a website, also called web developer, site author or website administrator), on the other hand, want to rank high in SERP for select-ed keywords [4] because high position brings them more traffic and hopefully more conversions [14]. Search engines don't reveal their algorithms, only general guides on how webpages should be constructed to rank high in their SERP [2, 5]. This process of fine tuning of websites to comfort to search engine guidelines with goal to rank higher in SERP is called search engine optimization (SEO). There are many factors that SEO includes [10, 15], and page title is one of them. In this paper we consider page title as the title specified in tag "title" in header of HTML (Hyper Text Markup Language - standard markup language used to create web pages) file.

Another important ranking factor is the number of backlinks a page has. Backlinks are incoming links to a website or web page, received by a web node (web page, directory, website, or top level domain) from another web node. Other web pages use various link texts to describe the linked page. In most cases this anchor text contain relevant keywords that search engines use as important ranking signal. In this paper we explore the possibility to use this anchor text from links to update and improve page title of targeted page. By inspecting DMOZ open directory (https://www.dmoz.org/) we found that 45.5% of webpages don't have SEO optimized page title, but do have good backlinks with relevant anchor texts. Webmasters or machines (software) could use this valued data to improve web page titles and gain more visibility in search engine results pages. There is also a big amount of web pages with poor titles that don't tell the user anything about the page (e.g. "untitled document" or "new page 1").

2 Background

SEO process can be divided in two groups: "on-page SEO" and "off-page SEO", depending on where we apply SEO techniques [10].

2.1 "On-page" search engine optimization

"On-page" or "on-site" SEO represents process of improving web page's content and HTML code to comply with search engine guidelines. Several research has been done to reveal what are most important "on-page" factors [15]. Results show that keywords should be used in following places [11]:

- Page title – in HTML tag "<title>" in page header. Text in this tag is displayed in browser title bar and in SERP as link to the page. It's important to have good title as this is the first thing a user sees in SERP and based on them will decide to click or not.
- Page description – in HTML tag "<meta>" in page header. This text usually in-cludes a few sentences of descriptive text of what is

this page about. Search en-gines use this text to display it as short description of webpage in SERP immedi-ately bellow page title.

- Domain name and URL. Search engines give extra scores to pages with keywords in domain name and URL. If webmaster can't affect the domain name, the rest of URL, folder and file names can for sure.
- Heading tags – in HTML tags "<h1>", "<h2>" and "<h3>". Web pages should use heading tags to represent titles in page's body. Having keywords in these tags is one of important ranking signal.
- Page body – text in HTML tag "<p>". Text on page with good keyword density is valued in SEO.
- There are many other "on-page" factors like site speed, code vs. content ratio, etc., but above listed represent the most important ones.

2.2 "Off-page" search engine optimization

While "on-page" SEO depends on webmasters knowledge of SEO best practices that any webmaster can learn and apply, "off-page" SEO is not quite under webmaster's control. In general "off-page" SEO is all about links. The more backlinks your website have, the more important it is in search engine eyes. It is not the count of backlinks that matters, but also the quality of these links – links from quality and authority websites counts more then links from small irrelevant sites. The most important algo-rithm and backlink metric in web information retrieval is Google PageRank, invented by Google founders and incorporated in Google search engine [6, 9]. To gain more backlinks, webmaster can apply his web page in web directories, post messages in relevant forums, use social media, etc., but most important links are natural links, created by users, spreading the good word about the website.

2.3 Anchor text

By anchor text we assume the text of the link from source page to target page. In HTML anchor text is constructed with tag "a", e.g.:

Social network

In the example above of the link that leads to "www.facebook.com" the anchor text is "Social network". Anchor text is used by search engine as important ranking factor. If many pages exists with above link to "Facebook" then search algorithm will deter-mine that "Facebook" is a "social network" – because a lot of other independent web sites are

telling so. When a user types keyword "social network" in a search engine, the search engine wants to satisfy user query and list most relevant results and rank them according to relevance. One of the important signal in this process is anchor text from backlinks that is a backbone of many today's search algorithm like Google's PageRank [9].

3 Related work

Similar research of taking advantage of links and anchor text has been done in [7]. Authors are proposing a new method – using anchor text – to automatically determine web page title for the purpose of organization and categorization of web pages. Besides using backlinks, there are many other methods that have been researched for automatically determine the web page title. These methods mostly rely on content analysis of the web page [13], HTML and CSS formatting inspection, DOM tree exploration, features extraction and linguistic information. Using backlinks information is different as it does not rely on content of the website. In [7] authors are using all links of targeted page by creating hyperlink graph with HITS algorithm and URL similarity using Levenshtein distance to measure and calculate the appropriate web page title. Levenshtein distance is a string metric for measuring the difference between two words calculated as the minimum number of single-character edits (i.e. insertions, deletions or substitutions) required to change one word into the other. The results of this methods shows that anchor text can be used to determine the web page title for categorization purpose. However, the objectives of this research are different then this paper – the results are not tested according to SEO rules and best practices.

Anchor text and his features have been main focus in [3]. Authors of this paper claim how anchor text behaves much alike user query and are exploring how is it related to document to gain better understanding of this relationship. According to [8] web page titles, anchor text and user query are related because they are constructed by the same mental process: users want to describe the topic (of the web page, search query or the document) in as few words as possible. In [3] authors are conducting an experiment of web search by page titles and web search by anchor text only. The results show that documents returned by matches on anchor text are more relevant then the one returned by page title. This is giving us a motivation that anchor text could be used to improve page title (in terms of SEO) of existing web pages with no title, bogus or irrelevant titles.

Similar research is done in [12] where authors are exploring how anchor text can be used to improve search quality. They are comparing various search techniques obtained from anchor text and document text, and how can these methods be fused to improve

ranking. Positive results from this research are also contributing to our idea of using anchor text in title construction.

4 Research methodology and results

To explore the usage of anchor text in order to improve web page titles in terms of SEO we used popular open directory DMOZ. We randomly extracted a sample of 1000 URLs from DMOZ and for each of them extracted top 5 backlinks and their anchor text. We used first anchor text that complied it with the no bogus anchor text and no URL rule (URLs in anchor text). For extracting backlink data we used Majestic SEO API (http://developer-support.majestic.com/api/). After processing all URLs we categorized URLs in several categories based on title and anchor text features as shown in Table 1.

Table 1. Categorization of inspected sample of web pages from DMOZ

Category	Description	Count
A	exact match of title and anchor text	125
B	title starts with anchor text	175
C	anchor text is found in title before position 40	108
D	similar match with Levenshtein similarity <0.5	92
E	Page title and anchor text don't match	399
F	no title or generic title, but anchor texts exists	56
G	no anchor texts	27
H	page does not exists	13
X	no title and no anchor text	5
Total:		1000

From statistics from Table 1 we can see that our method of adding top anchor text in page title can be used on categories E and F which represents 45.5% web pages in our sample. Categories A, B, C and D represent web pages with good page title (50%), while the rest, categories G, H and X could not be used in our experiment because of missing backlinks with generic anchor texts or other problems. Table 2 lists generic and bogus page titles and anchor texts that we found in our sample.

Table 2. Generic page title and top anchor texts found in web pages from DMOZ sample

Generic page titles	Generic anchor texts
Home, startseite, 0000, home page, index, untitled document, we've moved, Start, Unbenanntes Dokument, under construction, Document sans nom, inicio, Willkommen, welcome, uskoro, Documento sin título, Sin título, coming soon page, untitled	click here to view website, open in a new browser window, website, view website, click here to visit website, view site, untitled document, official site, official website, view my listings, web site, read more, unofficial homepage, company website, inicio, online shop, website öffnen, visit the site, address, read sources, more details, here, read more, about, continue reading, my personal website

Generic anchor texts represent bad texts to use in page titles so these texts were expelled from our method. Pages with generic page titles are categorized in group F and in our method replaced with top anchor text. After replacing (for category F) or adding top anchor text on the beginning of the page title (for category E) for all 455 websites we inspected the new page title of each web page to check if it relates to website topic. We used a software to check if selected top anchor text exists in page body, which was the case in 31% of websites, and we additionally manually checked the remaining 69% of websites to see if the relation between anchor text and website topic exists. In total we found that 81% of page titles constructed with our method were related to website topic and represented improvement in page title quality against previous state.

5 Conclusion and further work

Our experiment showed that 45.5% of inspected websites from DMOZ have bad page title according to SEO guidelines and that these website's page title could be improved if we use top anchor text extracted from backlinks of inspected website (in 81% cases). This method can be used only on existing web pages that have backlinks and can be part of a software for improving SEOs web pages. The limitation of this method is that it does not take into account the backlink quantity and quality: a web page with just 5 backlinks can give misleading top anchor text, or an anchor text from quality web page could be more accurate then the top anchor text where we observe only the quantity. Also, link bombing - link manipulation technique that attackers use to boost ranking for unrelated keywords [1], can disrupt our method and create misleading page titles. These limitation represent guidelines and challenges for further research in this field. This research used DMOZ as a source of websites, but it could be

interesting to inspect other available datasets for comparison of the results.

References

[1] Adalı, S; Liu, T; Magdon-Ismail, M. An analysis of optimal link bombs. *Theoretical Computer Science*, 437:1-20, 2012.

[2] Bing webmaster guidelines, http://www.bing.com/webmaster/help/webmaster-guidelines-30fba23a, downloaded May 10 2015.

[3] Eiron, N; McCurley, K. S. Analysis of anchor text for web search. In: *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, pages 459-460, Toronto, Canada, 2003.

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

[5] 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 10 2015.

[6] Vryniotis V. Is Google PageRank still important in Seach Engine Optimization?, http://www.webseoanalytics.com/blog/is-google-pagerank-still-important-in-seach-engine-optimization/, downloaded May 10 2015.

[7] Jeong, O. R; Oh, J; Kim, D. J; Lyu, H; Kim, W. *Determining the titles of Web pages using anchor text and link analysis*. Expert Systems with Applications, 41(9):4322-4329, 2014

[8] Jin, R; Hauptmann, A. G; Zhai, C. X; Language model for information retrieval. In: *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 42-48, Tampere, Finland, 2002

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

[10] Search Engine Land's Guide To SEO, http://searchengineland.com/guide/seo, downloaded May 10 2015.

[11] Su, A. J; Hu, Y. C; Kuzmanovic, A; Koh, C. K. How to Improve Your Google Rank-ing: Myths and Reality. In *Web Intelligence*, pages 50-57, Toronto, Canada, 2010.

[12] Wu, M; Hawking, D; Turpin, A; Scholer, F. *Using anchor text for homepage and topic distillation search tasks*. Journal of the American Society for Information Science and Technology, 63(6):1235-1255, 2012.

[13] Xue, Y; Hu, Y; Xin, G; Song, R; Shi, S; Cao, Y; Li, H. *Web page title extraction and its application*. Information processing & management, 43(5):1332-1347, 2007.

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

[15] 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.