Research online Publishing www.ronpub.com

Open Access

Open Journal of Web Technologies (OJWT)
Volume 5, Issue 1, 2018
http://www.ronpub.com/ojwt
ISSN 2199-188X

Webpage Ranking Analysis of Various Search Engines with Special Focus on Country-Specific Search

Sinan Babayigit, Sven Groppe

Institute of Information Systems (IFIS), University of Lübeck, Ratzeburger Allee 160, 23562 Lübeck, Germany,
sinan.babayigit@student.uni-luebeck.de, groppe@ifis.uni-luebeck.de

Abstract

In order to attract many visitors to their own website, it is extremely important for website developers that their webpage is one of the best ranked webpages of search engines. As a rule, search engine operators do not disclose their exact ranking algorithm, so that website developers usually have only vague ideas about which measures have particularly positive influences on the webpage ranking. Conversely, we ask the question: "What are the properties of the best ranked webpages?" For this purpose, we perform a detailed analysis, in which we compare the properties of the best ranked webpages with the worse ranked webpages. Furthermore, we compare country-specific differences.

Type of Paper and Keywords

Regular research paper: *search engine optimization, webpage ranking, ranking algorithms, country-specific search*

1 Introduction and Motivation

Today, everyone who wants to be found as an entrepreneur has to be present on the web. Since search engines are the most frequently used applications on the web, it is essential to be found among their search results and, preferably, to appear as best ranked as possible in those results. With a market share of over 94.5%, Google takes the position of the top dog among the search engines in Germany. If you take a look at Bing with a market share of about 4.2% and Yahoo with about 1%, you get a total coverage of all search queries of over 99% in Germany [28].

This paper is accepted at the *International Workshop on Web Data Processing & Reasoning (WDPAR 2018)* in conjunction with the 41st German Conference on Artificial Intelligence (KI) in Berlin, Germany. The proceedings of WDPAR@KI 2018 are published in the Open Journal of Web Technologies (OJWT) as special issue.

Nowadays most people search for everything imaginable several times a day, but they only look at the first page of search results. And on this page, also referred to as SERP (Search Engine Results Page), most users just click on one or more links of the top 5 results [26]. This illustrates that it is essential to be as best ranked as possible on the first SERP for the corresponding search queries.

Hence no one who wants to be found on the Internet with his website can ignore search engine optimization (SEO). Only with continuous search engine optimization website developers succeed to be permanently ranked on the first page of the search results for the relevant search terms. These relevant search terms or keywords describe search queries that users of a website typically search for.

In the context of this paper, the topic is widened and no specific sites are optimized. The respective top 10

keywords of the year 2016 are used as search queries for four different countries. For these keywords, the webpages are analyzed, which appear on the results pages of the search engines. The analysis includes various webpage properties such as the page title, the page load speed, or the number of words on the page.

For the analysis, a web-based tool was developed. In order to analyze country-specific differences, four countries (two developed, one emerging and one developing country) are considered to cover the full range from rich to poor. Therefore it is possible to take a cautious look whether or not there are any differences in the quality of webpages.

The contributions of this paper are hence a comprehensive analysis of various properties of best ranked webpages (by three different major search engines) including the analysis of differences in mobile and desktop search as well as examining country-specific search results.

2 SEARCH ENGINES

This section describes how search engines operate in general. We will also take a look at major search engine algorithms and the structure of the Search Engine Results Pages.

2.1 Crawling, Indexing and Searching

Search engines are foundational for any kind of search engine optimization. A search engine allows you to search for information. For this purpose search engines are creating an index. This happens in the background without the knowledge of the user. The search engine responds to search queries entered by the user in the form of search words (even multiple or complete sentences).

Figure 1 shows schematically the process of indexing a webpage:

1. A web crawler, also called bot or robot, searches the web to find HTML pages.
2. The web crawler stores these pages unchanged in its search engine's document database.
3. The pages now go through processes such as the removal of HTML tags and stop words. Stop words are words that are common in one language and have hence not much informative value for the page content.
4. The search engine now creates indexes by generating different representations of a page. These include, for example, single words or paragraphs. Furthermore, information about incoming and outgoing links is noted and a snippet is generated.
5. The search engine stores the indexes in its index database.

Figure 2 shows the schematic concept of performing a search request with subsequent delivery of the results:

1. The user enters a search query, typically a single word or a short sentence. As in step 3 of the indexing process stop words are removed here, too.
2. The search engine generates an ordered list of URLs that it considers relevant based on the index database. The search engine then displays the snippets in the SERPs that correspond to the listed URLs.
3. The user clicks on certain snippets to access those webpages.
4. The retrieval-evaluation component of the search engine collects feedback about the relevance of the websites. This is usually done by clicking on relevant links (CTR) but also via direct feedback.
5. The relevance feedback is used to improve the SERPs of the search query for the user. The search engine could rewrite the search query and run it again. This process repeats until the user is satisfied with the search results or the session is terminated.
6. The search engine stores meta information such as search queries, direct relevance feedback and indirect feedback about the clicked snippets in the logging database. These data are used to improve the search in the future.

2.2 Main Search Engine Algorithms

Search engine algorithms are used to rank indexed pages according to the searched keywords. In the following sections, we will take a brief look on two well-known search engine algorithms and then look at the algorithms used today.

2.2.1 PageRank Algorithm

The PageRank algorithm [24] is a method of assigning a numeric weight to a webpage. This weight represents the determined importance of each page. To determine this importance, the algorithm considers the backlinks that point to a page as well as the so far determined importance of the pages containing the backlinks, i.e., the weight of a page increases as the page receives more backlinks from more important pages. As a matter of fact, the PageRank algorithm first initializes each page with the same weight and recomputes each weight in several iterations (until the weight converges). Higher weighted pages are considered as the most important ones with a high popularity and have a good chance of appearing on the first SERP. The algorithm was developed by Larry Page and Sergey Brin, the founders

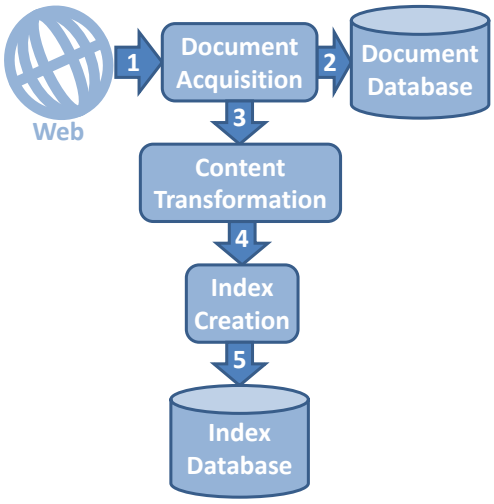

Figure 1: Schematic representation of an indexing process (adapted from [13])

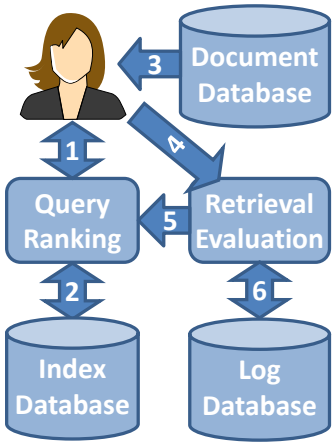

Figure 2: Depiction of a search process (adapted from [13])

of Google, at Stanford University, who applied for a patent for the PageRank algorithm.

2.2.2 Hilltop Algorithm

The Hilltop algorithm [3] is, referring to search engines, a method that sorts a large number of linked webpages according to their relevance to the search queries entered by the user. In essence, the hilltop algorithm consists of two consecutive phases. First, the so-called expert pages of a keyword are determined, which can be done automatically. Expert pages are defined in the Hilltop algorithm as follows: An expert page refers to a large number of independent other pages that also match the search term or topic. After this first phase, the authority pages are sorted based on the frequency and relevance of the referring expert pages. By definition, authority pages are webpages that refer to at least two independent expert pages. This procedure can be problematic if no expert pages are available for a search term. Then the Hilltop algorithm can not provide a result. The algorithm was developed by Krishna Bharat and George A. Mihaila at the University of Toronto.

2.2.3 Today's used Algorithms

Today, many features of the above and other algorithms are used in combination, as far as can be concluded from all known SEO factors. According to the known SEO factors, the score of a webpage consists of the search words that appear in the title, meta tags, headings, the content of the page, etc., and the backlinks [25]. It should be noted that due to the nondisclosure of the algorithms not all properties are known that affect the score of a webpage, and thus their ranking. There are however reasonable assumptions about which other properties could be relevant. Dean [5] released a list of 200 possible factors that could affect the ranking of webpages. Some of these are proven, others controversial and some highly speculative. A newer technology, which could also be a possible factor for webpage ranking, is Accelerated Mobile Pages (AMP)[1]. This is mainly used by Google for its mobile search and promises fast loading webpages on mobile devices. The influence of AMP on the ranking of webpages is also examined in this paper.

2.3 Structure of Search Engine Results Pages (SERPs)

The pages on which the result links of a search query are presented are called Search Engine Results Pages or shortened SERPs. The SERPs of the search engines that were examined in the context of this paper contain three types of search results. These are the organic search results, advertising for promotional purposes and the latest news. In the following, each group will be explained in more detail.

2.3.1 Advertising for Promotional Purposes

In addition to the other search results, paid content called ads are displayed on the SERPs for advertising purposes. The ads are usually displayed in a highlighted area and labeled as advertisements. Such offerings for ads are called Google AdWords[2], Bing Ads[3] and Yahoo Search

[1] https://www.ampproject.org/
[2] https://adwords.google.com/home/
[3] https://secure.bingads.microsoft.com/

Marketing[4].

2.3.2 Latest News

Also latest news are displayed on the SERPs in addition to the other search results. However, in order to be considered as latest news page, a website has to apply to the search engines and there is a manual review of the website.

2.3.3 Organic Search Results

The natural search results of a search engine are called organic search results. These are ranked according to the previously discussed aspects of search engine optimization and content of good quality. Although a direct automatic detection of good quality content is difficult or even impossible for a search engine, backlinks in mere numbers or from well-known websites are always an indicator of good content quality. Since only the organic search results can be influenced in their ranking by using search engine optimization, only these organic results and their ranking are analyzed within the framework of this paper.

3 Search Engine Optimization

According to Google [11] SEO covers various modifications of websites and techniques making it easier for search engines to crawl, index and understand a website. There are two types of SEO: On-page optimizations deal with modifications of the website structure. Off-page optimizations describe techniques being independent of the website structure. Considering the right combination of both can lead to significant better positions in the SERPs. Furthermore, SEO techniques are typically split into White-Hat-SEO, which contain techniques for a good design of websites, Black-Hat-SEO utilizing bad designs of websites for better rankings [19], and Grey-Hat-SEO covering all other techniques between these extremes.

3.1 Black-Hat-SEO

Not all web designers adhere to proven methods of search engine optimization. Unethical practices like the generation of spam, which affect the ranking quality, occur more often in recent years. Indeed this kind of practices combined with the increasing importance of the Internet as a market instrument led to the creation of a new type of business being specialized to improve the position of their clients in SERPs. Although most web designers of this business favor the proven methods of SEO, i.e., White-Hat-SEO, some apply the dubious practices of Black-Hat-SEO for achieving a good ranking. This search engine spam is a serious problem with estimated costs of 130 billion USD [22] for those companies suffering from a worse ranking caused by Black-Hat-SEO of their competitors.

Spam of search engines deals with various different fraudulent techniques for achieving a top ranking in the first SERP: For the purpose of improving the ranking websites are created fooling the algorithms of search engines. Because of search engine spam a webpage without relation to the search query may be even ranked quite well in the search results. This kind of techniques lowers both the efficiency and the correctness of search engine results. Hence search engine providers take actions to face problematic search engine spam. For example, in the early days of search engine optimization meta keywords defined in meta tags of website headers were the basis of indexing a website. A web designer could use meta keywords in order to specify keywords, which are not necessarily relevant to the content of the webpage. In order to avoid this misuse, major search engines do not support meta keywords since 2009 any more. However, some website providers still use meta keywords as before. Keyword stuffing describes in this context the unnatural repetition of certain keywords for the purpose of increasing the frequency of these keywords. Keyword stuffing is today even punished by major search engines resulting in a worse ranking.

Current practices of search engine spam cover automatically generated page content, cloaking, deceitful redirects, link spam, doorway pages, hidden text and links, spam from affiliate programs as well as embedded malicious behavior [13].

3.2 White-Hat-SEO

Search engine providers offer starter guides of search engine optimization and web master tools to combat search engine spam and to help web designers to develop websites according to White-Hat techniques. Starter guides provide rough ideas about the way search engines index documents and process search queries without disclosing secrets and detailed strategies of search engines. Webmaster tools check if recommendations of White-Hat-SEO are followed.

The search engine Google is an example for how search engines establish their leadership in developing best practices for the Internet. Google's Panda update came in the year 2011. Panda periodically applied a filter on the whole index of Google decreasing the ranking of websites with bad quality and increasing the ranking of high-quality websites. Google's Penguin update in the year 2012 further cut search engine spam:

[4] https://advertising.yahoo.com/

Websites distributing spam links received afterwards a lower ranking or were even deleted. The Hummingbird update in the year 2013 ranks websites according to the semantics of websites. Hummingbird unites the purpose and the contextual meaning of a search engine being more effective than an algorithm based on pure keyword frequency. Google's Pigeon update published in the year 2014 considers contexts like the current location of the user to rank local information in a better way. Google's Mobile update also known as Mobilegeddon in the year 2015 favors user-friendly websites for mobile users. In the year 2016 Google applied another mobile update, which further improves the ranking of responsive websites for mobile search. In the beginning of the year 2017 Google updated its algorithms to punish websites with aggressive pop-ups for advertisement purposes, which affect user experiences. [21] In July 2018, Google rolled out the Speed Update for all mobile users, where the page speed is now also a ranking factor for mobile searches [12] (which has been a ranking factor for desktop searches already for a long time).

White-Hat-SEO embraces two main categories: The first category is on-page optimization dealing with the structure and content of a website. On-page optimization covers techniques from simple diction to mechanisms for limiting the indexing of search engines.

The second main category of White-Hat-SEO is off-page optimization dealing with best practices for embedding incoming and outgoing external links. A carefully designed directory structure not only helps the administration of websites but also bots for the purpose of indexing the whole website. The navigation structure of a website range from breadcrumb navigation (displaying the user the current branch of the application) to site maps (containing the website structure). These types of navigation are intuitive and quite natural for human users as well as for web crawlers. On the contrary, drop-down menus generated by JavaScript code might be problematic for web crawlers. For a good indexing all links should be provided in textual and not in graphical form. Navigation structures also include meta tags for the web crawler for controlling the indexing on site level.

3.3 Further Related Work

Bifet, Castillo, Chirita and Weber [4] approximate ranking algorithms by analyzing the SERPs. First, a numeric value is determined for each query, describing the observed webpage properties and transforming each document into a vector. Then they look at the differences of these vectors in relation to their ranking. Vectors that are more likely to be in the top direction receive a "+" and vectors that reflect rather poor search results get a "-". Hereby, the original problem becomes one of binary classification. When two vectors (of documents) are taken they try to predict which document will be ranked better. In contrast to our analysis, more attention is paid to the occurrence of the search term in all possible parts of the document. We limit ourselves only to the occurrence of the search term in the page title, but also examine webpage properties from completely different areas. Here, in addition to the occurrence of the search term they also take a look at the backlinks which is done in the context of our paper, too.

Evans [8] is looking for the most popular techniques to rank best on Google. For this purpose, the results of a study in which 50 highly optimized websites are created as part of a contest are presented. In the context of this contest they examine the PageRank, the number of pages of a domain, the domain age, the backlinks and the use of third party sites such as directories and social bookmarking sites. A comparative study is also conducted for 50 non-optimized sites. The study concludes that PageRank and presence in directories and social bookmarking sites is important. The backlinks are also very important, which we also examine in the context of our paper. In contrast to the approach shown here, we consider more webpage properties and analyze in more depth. Furthermore, we do not use specially optimized webpages as references.

Bar-Ilan [2] investigates how similar rankings behave on different search engines when the results overlap (i.e. occur in both search results). The ranking of search results for identical search queries is analyzed. The search engines used are Google, AlltheWeb, AltaVista and HotBot. To compare the similarity of the rankings of two different search engines, the rank correlation coefficient is calculated. On the other hand, if more than two rankings are compared, Kendall's coefficient of concordance [17] is calculated. For the analysis, 15 queries are submitted to the four search engines returning a disjoint set of 16985 results. Unlike our approach, these result links are not analyzed any further in detail. However, they are merely compared to each other as a whole according to the methods described above. Indeed the results of this analysis show that the search engines use very different ranking algorithms. A circumstance, which we observe as well in relation to some website features, if not for all.

Gandour and Regolini [9] describe methods that have been used to optimize a website[5]. Mainly due to on-page optimization, they observe rapid improvements in the ranking of the website. The improvements include title tags, the visible text as a whole, the name of the URL, meta tags and images "alt" and "title" attributes. From this perspective, webpage properties are examined that

[5] www.fragfornet.grenoble.cemagref.fr/

are not considered in our analysis. This also applies vice versa. The authors acknowledge that there has not yet been any off-page optimization of the backlinks, which is generally considered to be a key issue. In contrast to our analysis, however, their contribution covers a specific attempt to optimize an existing website. Our approach is more general and not aimed at a specific website.

Mavridis and Symeonidis [18] look at website properties in a Web 2.0 and Web 3.0 context. Among other things, they examine the influence of metrics on the ranking, where the metrics take semantic data into account. It turned out that this has a noticeable impact on the ranking, especially on Bing, but also on Google. Only Yahoo seems to attach little importance to it. This seems surprising as Dean [6] specifically explore the use of schema markup and found no discernible impact on the rankings of Google results. Since Dean's results are extensive and new, our analysis does not include a study of metrics that take semantic data into account. However, they may gain more importance in the future.

Bar-Ilan, Mat-Hassan and Levene [1] compare rankings of search results over periods of time. For this purpose, five searches are observed daily for 21 days. This is done for a text search with the search engines Google, Yahoo and Teoma and for an image search with Google, Yahoo and Picsearch. After about three months, the procedure is repeated and the results and rankings of the two periods are compared to each other. The authors observed that the image search results are not as stable as the text search results. Furthermore, there are hardly any overlaps in image search results when different search engines are used. In text search, these overlaps are much higher. Hence the authors conclude that either the ranking algorithms for the text search as opposed to the image search are more similar among the search engines, or the image databases of the various search engines are almost disjoint. The latter seems more likely if you include Bar-Ilan's study above and our findings. In contrast to our analysis, however, the image search is still considered here. We restrict ourselves only to text search, but for desktop and mobile devices.

Zhang and Dimitroff [33] analyze the impact of metadata on website ranking. The authors execute an experimental study, in which metadata turned out to be important. The results suggest that metadata is a great way to increase the visibility of a webpage in prospect of search engines. Hence the meta tag with name = "subject" plays an essential role. We do not analyze the use of meta tags. In particular, the meta tag with name = "keywords" was much discussed and sometimes misused, so that Google completely renounces its use today [10].

Su, Hu, Kuzmanovic and Koh [30] look at how to improve Google's ranking. For this purpose, they concentrated on the ranking algorithm and tried to recreate it. The authors claim that Google's ranking algorithm can be replicated relatively accurately using reverse engineering. This is a completely different approach than ours. However, the authors also examine whether and acknowledge that the presence of the keyword in the page title is important. Our results do not indicate that. Furthermore, the authors examine the importance of PageRank, which is stated to be very important. Regarding this aspect, we come to the same result, even though we have not considered the PageRank as such but only the backlinks. Other significant findings of the authors concern the HTML syntax and blogs. As a result syntax errors do not affect the ranking and Google is generally set rather negative in terms of blogs.

Egri and Bayrak [7] explore the importance of page load speed, a reduced bounce rate, the number of page views, the length of stay on the page, and the page layout to keep users on the site. Like us, the authors study the importance of page load speed. This is an important feature of websites. Our results confirm their result, especially in relation to the mobile search.

Apart from the very extensive, but only on Google focused, analysis by Dean [6], we did not find to the best of our knowledge, other work that deals with the use of HTTPS, responsive web design and especially the newer AMP documents. Furthermore, we have not found any comparable analysis that considers the impact of the number of words on a webpage and the length of an URL in terms of ranking. All this is evaluated not only for the three major search engines Google, Bing and Yahoo, but also for different countries and separated by mobile and desktop search results.

4 Experimental Analysis

In order to be able to analyze the properties that are responsible for a very good or rather weak ranking in the SERPs, a tool is needed. This tool should extract the result links from the SERPs in order to be able to open and analyze them with the analysis functions developed in this paper.

4.1 Choice of Parameters

For the analysis of the webpage rankings, we need to choose the considered search engines, the corresponding search terms for which a ranking is examined, and a selection of webpage properties according to which a webpage is examined.

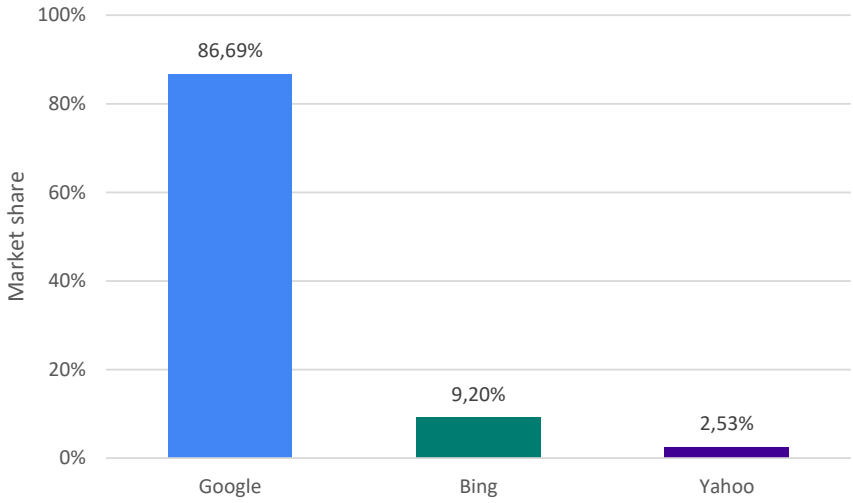

Figure 3: Desktop search market share of search engines in Germany [29]

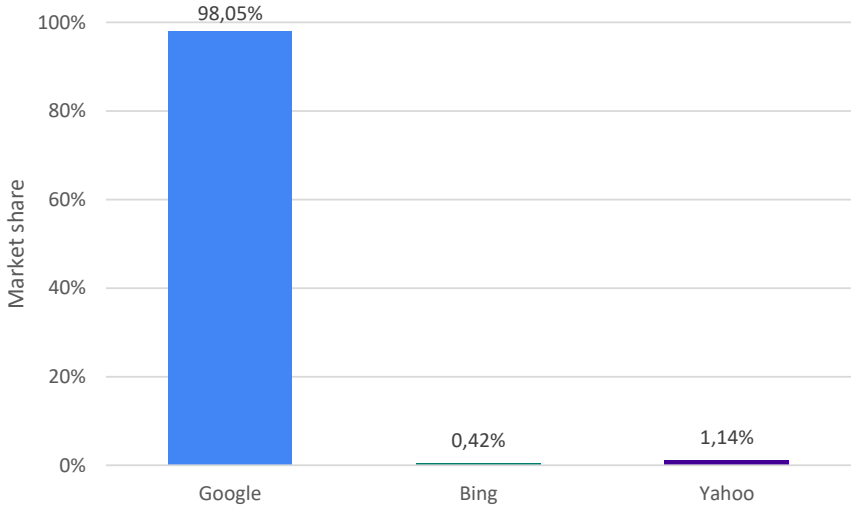

Figure 4: Mobile search market share of search engines in Germany [29]

4.1.1 Choice of Search Engines

The search engines considered in this paper are Google[6], Bing[7] and Yahoo[8]. These are the three largest search engines in Germany with a total market share of nearly 99% [29] for both the desktop, see Figure 3, and for mobile search, see Figure 4.

4.1.2 Choice of Analyzed States

We need to determine a selection of countries, which are particularly suitable to get a rough picture of whether there are differences between countries in the website quality. In order to further increase the significance, we consider differences in terms of wealth and general quality of life. The result is a small group of four countries:

- Federal Republic of Germany
- United States of America
- Republic of India
- Central African Republic

The Federal Republic of Germany is selected for obvious reasons. This paper is written at a German university and therefore it is useful to take a look at the website quality here in this country. Furthermore, the Federal Republic of Germany is one of the industrial nations and thus one of the richest countries in the world.

The United States of America are chosen because they are the home of the three search engines analyzed in this paper. The United States of America has great

[6] https://www.google.com
[7] https://www.bing.com
[8] https://www.yahoo.com

importance in the world and represents a significant economic power. Like the Federal Republic of Germany, the United States of America belongs to the group of industrial nations.

The Republic of India is, after the People's Republic of China, the second most populated country in the world with over 1.3 billion inhabitants (in the year 2016) [23]. Thus, it is the most populated democracy in the world. India is particularly interesting for this work because it belongs to the group of emerging economies [14] and thus a possible difference to websites from industrial nations can be analyzed.

The Central African Republic is a highly underdeveloped, unstable state. In terms of gross domestic product per capita, the country was the poorest in the world in 2016 [15]. In the Human Development Report of 2016 it was in last place of 188 countries [31]. Thus the country generally represents a strong contrast to the industrial nations and is therefore well suited to investigate possible differences in website quality in comparison to emerging and industrialized countries.

4.1.3 Choice of Keywords

As this paper also aims to analyze possible country-specific differences in website quality, we used the Google Trends[9] of the year 2016 to refer to the respective top 10 keywords of the surveyed countries. The investigation with these search terms is not limited to Google. These search terms are also used for the searches with Yahoo and Bing, since these search engines themselves do not provide a country-specific search term top list [27, 32].

The exception is the Central African Republic. This country is not included in Google's top charts. Thus, the general trends[10] of 2016 are used. They are very similar to the top charts. For reasons of consistency, the top 10 are considered here, too.

Table 1 contains the respective top 10 keywords of the year 2016 for the countries Germany, USA, India and the Central African Republic.

4.1.4 Choice of Analyzed Webpage Properties

As the title of this paper suggests, webpages are compared with each other on the basis of special characteristics, in order to get a picture of which properties have particularly high or very little influence on the webpage ranking on search engines.

[9] `https://trends.google.de/trends/topcharts` (all retrieved on 4 September 2017)

[10] `https://trends.google.de/trends/explore` (retrieved on 4 September 2017)

The functions for analyzing webpage properties realized in this paper are described in more detail below. Here it should be noted that the focus of this paper is to provide a good range of analysis functions available to get more meaningful results from the analysis later. The functions are therefore not optimized to the last detail, or may be faulty. However, this applies only to a few cases, which are mentioned in the more detailed description of the respective function. The analyzed webpage properties are at a glance:

1. Search contains title of webpage
2. Number of words on webpage
3. Webpage is encrypted
4. Webpage is responsive
5. Webpage is an AMP document
6. Length of the URL of webpage
7. Load time of webpage
8. Number of backlinks to the webpage

4.2 Conception of the Analysis Tool

We describe the implementation details of our analysis tool in this section, especially how to retrieve and determine the properties of webpages and how to execute the different types of searches, i.e., desktop, mobile and country-specific searches.

4.2.1 Implementation Details of the Analysis Functions

We deal with the analysis functions for the chosen webpage properties in the following paragraphs:

Search contains title of webpage: In order to check whether the search is contained in the title of a webpage, the document object model (DOM) of the respective page is first loaded. Afterwards, the DOM is searched for the contents of the title tag. Since all searches and webpage titles are stored in a database, it is easy to check whether the search is included in the title of the webpage.

Number of words on webpage: Another important feature in terms of search engine ranking is the number of words on a webpage. According to Dean [6], the average search result on the first SERP of Google has 1890 words. As a result, longer content tends to rank better in Google SERPs. Therefore, in the context of this paper, we develop a function to count the total number of words on a webpage. The function is based on the HTML source code of the respective page by deleting all contents between opening and closing Script, Style and

Table 1: Top 10 keywords of the year 2016 of all considered countries

#	Germany	United States	India	Central African Republic
1	EM 2016	Powerball	Rio 2016 Olympics	اغاني (Arabic: music)
2	Pokemon Go	Prince	Pokémon GO	mhd
3	iPhone 7	Hurricane Matthew	EURO 2016	galopfrance
4	Brexit	Pokémon Go	Sultan	افلام (Arabic: movies)
5	Olympia	Slither.io	Kabali	euro 2016
6	Tamme Hanken	Olympics	iPhone 7	facebook se connecter
7	Dschungelcamp	David Bowie	Indian Premier League 2016	unops
8	Nico Rosberg	Trump	Donald Trump	kenya airways
9	David Bowie	Election	Udta Punjab	www.m facebook.com
10	Donald Trump	Hillary Clinton	P. V. Sindhu	facebook email

Head tags as well as all HTML comments. Afterwards, all HTML tags are removed. Finally only the relevant words are left, which can be counted with a simple function.

Webpage is encrypted: Since the major search engines confirmed HTTPS as a ranking signal, we are able to rely on the analysis of this property done by the search engines. Thus, we only check whether the URL of the respective result link starts with ”https” or only with ”http”.

Webpage is responsive: Responsive web design describes the ability of a webpage to respond to various features of the particular device used, especially smartphones and tablets. The graphical representation of a responsive webpage is based on the requirements of the particular device the webpage is viewed with. For example, the layout of the whole webpage and displaying individual elements such as navigation bars or page columns are adjusted according to the device. The technical basis for this are the newer web standards, especially CSS3 Media Queries. Media queries assign a CSS style sheet to an output medium. To do this, CSS defines @media rules that are executed only if certain conditions are met. Hence to check whether or not a page is responsive, all CSS style sheets are searched for occurrences of the word ”@media”. Our analysis tool hence needs to retrieve all CSS stylesheets. For this purpose, the DOM of the respective page is loaded. Next, our tool looks for external style sheets, loads them, and searches for the occurrence of the @media string. If nothing is found in these external style sheets, the content of all page-internal styles will be searched. If nothing is found there as well, the site is rated as *not responsive*.

Webpage is an AMP document: Accelerated Mobile Pages (AMP)[11] is a derivation of HTML developed especially for the creation of websites running on mobile devices such as smartphones, tablets, etc. AMP is a relatively new technology released in the fall of 2015 by the AMP Project, led by Google. There are many HTML elements which are not allowed in AMP documents, because AMP supports multiple own custom element types. Thus, it does not represent a subset of HTML. The MIME type (according to the Specification Multipurpose Internet Mail Extensions), also called the Internet Media Type or simply Content-Type, is like in other HTML documents text/html and the file extension is .html or .htm, respectively. The opening HTML element is marked additionally with a lightning symbol (Unicode: 0x26A1) or just with the string ”amp”. To analyze whether a page uses AMP, a relatively simple approach is chosen. First, the DOM of each page is loaded. It is then checked whether the opening HTML tag contains the attributes that are required for AMP. As mentioned earlier, these attributes are either the lightning symbol (Unicode: 0x26A1) or the simple string ”amp”.

Length of the URL of webpage: As already indicated above, all URLs are stored in a database. Therefore, the length of an URL can be determined using a simple database function to get the length of a string.

Load time of webpage: Another important webpage property is the load time of the page. Load time measurement is a process of measuring the time it takes for a webpage to fully load all of its content. In order to obtain reliable experimental results, several measurements of the load time are necessary, because sometimes a web server has to process many requests at the same time. In such a case, the loading process would take longer, but

[11] https://www.ampproject.org/

is more of an exception and does not reflect the actual time it takes to load a page. In the context of this paper it is therefore decided to measure three times. This is done using a headless browser, which is a kind of web browser without a graphical user interface. In order to be able to measure the load time of a page correctly, a browser is required that can load all contents of a webpage. As already mentioned, always three measurements of the load time per webpage are made and stored in a database. For an analysis, the arithmetic mean may be taken from these values. An alternative approach is the use of the median from the three values, however with this approach small differences are ignored. Therefore, we use the arithmetic mean, although not completely correct from a statistical point of view.

Number of backlinks to the webpage: This paper takes a look at two types of backlink data:

- **Link pop:** Link popularity describes the number of inbound links to a page.
- **Domain pop:** Domain popularity refers to the number of inbound links from different domains.

The website `https://www.seokicks.de/` currently has an index of more than 200 billion link records (February 2017). This seems like a drop in the bucket, considering that there are around 1.9 billion websites in the world right now [16]. It should be kept in mind that the number of webpages is much greater, and therefore the number of links between these webpages too. However, we decide to use the above source, as it is free to use and may provide insight in the analysis of the results later. For example, it is conceivable that there is a correlation between a good ranking in a search engine and the presence in the index of the 200 billion link records.

4.2.2 Procedure for Obtaining the Correct Search Results for different types of Searches

For retrieving the search results for mobile search (even when running on a desktop computer), the analysis tool just modified the user agent to simulate a mobile device.

In order to retrieve the search results of a given country, the analysis tool modifies the search request (i.e., the url to retrieve the search results), because search engines consider the user's location for ranking the search results per default. The manipulations of the search request is different for the considered search engines:

Google: For a correct country-specific search, the search request must use the parameters "top-level-domain", "hl" (host language) und "gl" (country code). Furthermore, our analysis tool additionally uses the parameter "pws=0" (personal web search), which disables the personalized search [20].

Bing: Country-specific results are determined by using the parameter "cc"(country code) when searching with Bing.

Yahoo: Yahoo provides country-specific sub-domains for country-specific searches.

4.3 Analysis

In this section the data collected in September 2017 by our tool developed in the context of this paper are examined. For each of the search engines and countries considered in this paper, the numbers determined by the tool are looked at in detail and analyzed. These determined numbers represent how large the percentage or the average of the corresponding data is, that fulfills a property. Hence you can delineate the corresponding data to make differences clearly visible. As a form of representation, tables are chosen which represent the well ranked and badly ranked search results from left to right. SERP refers to the respective results page, that contains 10 result links each, which seems to be the standard using the common search engines. In this way, for each country, the average of the first (1), second (2), and third (3) SERP that satisfies a property is considered. Then a mean is computed of these top 30 results, followed by a vertical double-dash, which is the demarcation to the average of the worst 30 results. Afterwards we list individually the third last ($n-2$), the second last ($n-1$) and the last (n) SERP. Thus, this form of representation shows the natural flow through the search results and allows a simple comparison of the top 30 results with the worst 30 results. As an abbreviation for the respective countries their top-level domains are used for reasons of clearness.

4.3.1 Google

This section analyzes the results of the Google search.

Search contains title of webpage: According to the results in Table 2 in general in the industrialized countries, Germany and the United States, search queries are more likely to appear in the title of the webpage, as opposed to the countries India and the Central African Republic. Nevertheless, it is noticeable that the difference between the best search results and the worst search results is not very big. This applies to both desktop and mobile search.

Table 2: Google: Search contains title of webpage [data in pct.]

SERP	1	2	3	∅	∅	$n-2$	$n-1$	n
US:								
Desktop	100	100	96	**99**	**99**	97	100	100
Mobile	100	100	94	**98**	**99**	96	100	100
DE:								
Desktop	99	90	90	**93**	**99**	96	100	100
Mobile	99	90	90	**93**	**95**	86	100	100
IN:								
Desktop	94	80	61	**78**	**77**	66	74	92
Mobile	93	81	61	**78**	**72**	62	72	83
CF:								
Desktop	71	51	40	**54**	**38**	30	34	49
Mobile	62	40	40	**47**	**39**	25	32	60

Table 3: Google: Average number of words [data in thousands]

SERP	1	2	3	∅	∅	$n-2$	$n-1$	n
US:								
Desktop	2.4	2.1	1.3	**1.9**	**2.1**	2.5	2.5	1.4
Mobile	2.6	1.7	0.8	**1.7**	**2.2**	1.4	2.2	2.9
DE:								
Desktop	2.4	1.6	1.0	**1.7**	**0.8**	0.7	1.0	0.8
Mobile	2.1	0.8	0.9	**1.2**	**0.7**	0.7	0.8	0.6
IN:								
Desktop	1.5	1.2	1.6	**1.4**	**1.2**	1.4	1.2	1.0
Mobile	1.5	1.1	1.5	**1.4**	**1.0**	1.2	1.0	0.8
CF:								
Desktop	0.9	1.1	0.9	**1.0**	**1.5**	1.8	1.1	1.6
Mobile	0.9	1.4	1.2	**1.2**	**2.0**	2.8	1.6	1.5

Average number of words: The data in Table 3 is very contradictory: In Germany and India, more words seem to indicate a better ranking, but it seems to be the other way round in the United States and the Central African Republic.

Webpage is encrypted: According to the data in Table 4 HTTPS encrypted webpages are better rated more frequently overall. Only the desktop search in Germany is an exception here, but this does not really seem significant, as Google has already confirmed HTTPS as a ranking signal.

Webpage is responsive: Table 5 contains data about how many webpages are using responsive web design. It is noticeable that for both the best ranked and worst ranked results, the proportion of webpages that are responsive is always larger for the desktop search (though sometimes minimal or equal). This is surprising since

Table 4: Google: Webpage is encrypted [data in pct.]

SERP	1	2	3	∅	∅	$n-2$	$n-1$	n
US:								
Desktop	66	41	44	**50**	**40**	39	44	37
Mobile	60	42	48	**50**	**37**	37	41	33
DE:								
Desktop	40	42	39	**40**	**47**	50	54	38
Mobile	48	43	43	**45**	**44**	44	44	43
IN:								
Desktop	52	41	40	**44**	**35**	37	40	27
Mobile	44	42	48	**45**	**34**	37	36	29
CF:								
Desktop	64	61	58	**61**	**60**	58	63	58
Mobile	71	59	53	**61**	**57**	62	55	54

Table 5: Google: Webpage is responsive [data in pct.]

SERP	1	2	3	∅	∅	$n-2$	$n-1$	n
US:								
Desktop	90	92	91	**91**	**86**	84	88	87
Mobile	78	87	92	**86**	**86**	90	85	83
DE:								
Desktop	92	92	85	**90**	**89**	90	90	88
Mobile	90	89	95	**91**	**80**	86	77	76
IN:								
Desktop	98	94	89	**94**	**75**	81	81	62
Mobile	92	92	85	**90**	**70**	75	78	57
CF:								
Desktop	86	81	85	**84**	**90**	93	90	88
Mobile	70	81	75	**75**	**83**	86	86	76

responsive web design is especially important on mobile devices because of their smaller screen sizes.

Webpage is an AMP document: Although AMP documents appear to be less common in poor countries such as the Central African Republic (see Table 6), there is, across all countries, a clear difference as one compares the well ranked results to the badly ranked results. Furthermore AMP documents are much more common in the top 30 results than in the last 30 results.

Average URL length: According to the results presented in Table 7 a shorter URL is desirable if a webpage should be better ranked. Only the Central African Republic is slightly out of line here, although the values are relatively balanced.

Average page load time: Worse placed webpages not always load slower than higher ranked pages (see Table 8). Of course, this could indicate less rich content but you can not conclude on the basis of the data that faster

Table 6: Google: Webpage is an AMP document [data in pct.]

SERP	1	2	3	∅	∅	$n-2$	$n-1$	n
US:								
Desktop	0	0	0	**0**	**0**	0	0	0
Mobile	19	21	18	**19**	**5**	6	6	2
DE:								
Desktop	0	0	0	**0**	**0**	0	0	0
Mobile	34	30	18	**27**	**4**	6	1	5
IN:								
Desktop	0	0	0	**0**	**0**	0	0	0
Mobile	19	23	20	**21**	**4**	6	5	0
CF:								
Desktop	0	0	0	**0**	**0**	0	0	0
Mobile	3	14	12	**10**	**5**	4	7	4

Table 7: Google: Average URL length

SERP	1	2	3	∅	∅	$n-2$	$n-1$	n
US:								
Desktop	61	59	61	**60**	**78**	77	74	82
Mobile	56	65	64	**62**	**78**	77	77	79
DE:								
Desktop	73	73	70	**72**	**84**	84	84	83
Mobile	77	75	70	**74**	**83**	91	82	76
IN:								
Desktop	62	69	63	**65**	**84**	82	86	83
Mobile	62	71	71	**68**	**78**	76	77	80
CF:								
Desktop	54	76	74	**68**	**70**	69	69	71
Mobile	59	80	97	**79**	**78**	72	74	88

loading webpages automatically rank better. However, what is striking is that the results of the mobile search, for both the best and the worst ranked results, always load faster than the results of the desktop search. This also applies unrestrictedly across national borders.

Number of backlinks: Table 9 contains the number of backlinks in the form of the link pop. Overall, a small majority of the better ranked webpages can be seen, which is not surprising, as this is a core property for the ranking of webpages done by search engines. That the numbers of backlinks are sometimes so close together is probably due to the used source of backlinks, which does not appear to be comprehensive enough. These results, however, are slightly modified when looking at Table 10. Here, the backlinks are in the form of the domain pop which is much more meaningful. The data in Table 10 indicates that only the mobile search in Germany is an outlier.

Table 8: Google: Average page load time [data in seconds]

SERP	1	2	3	∅	∅	$n-2$	$n-1$	n
US:								
Desktop	5.2	5.8	6.2	**5.7**	**5.2**	5.1	5.3	5.4
Mobile	3.5	4.0	3.8	**3.8**	**4.6**	4.3	4.4	5.1
DE:								
Desktop	4.5	4.8	5.2	**4.8**	**4.1**	4.0	4.1	4.2
Mobile	2.3	2.9	3.5	**2.9**	**2.8**	3.4	2.6	2.3
IN:								
Desktop	4.6	7.1	6.3	**6.0**	**4.6**	5.0	4.6	4.2
Mobile	3.4	4.0	3.9	**3.8**	**4.5**	4.7	4.6	4.2
CF:								
Desktop	3.4	4.1	4.0	**3.8**	**4.6**	4.3	4.9	4.5
Mobile	3.0	2.9	3.6	**3.1**	**3.7**	4.0	3.4	3.6

Table 9: Google: Number of backlinks (link pop) [data in thousands]

SERP	1	2	3	∅	∅	$n-2$	$n-1$	n
US:								
Desktop	27.2	10	7.5	**14.9**	**0.2**	0.2	0.1	0.5
Mobile	22.2	4.8	0.5	**9.2**	**2.3**	3.2	3.6	0.1
DE:								
Desktop	6.2	0.3	0.6	**2.4**	**4.0**	0.0	0.0	12.0
Mobile	2.6	0.3	0.1	**1.0**	**58.8**	0.0	168	7.7
IN:								
Desktop	12.3	1.7	1.5	**5.2**	**1.3**	0.2	0.1	3.5
Mobile	2.5	0.3	0.2	**1.0**	**3.1**	0.0	0.0	9.2
CF:								
Desktop	793	46	1021	**620**	**31.9**	53	29	14
Mobile	918	7.8	0.0	**309**	**10.0**	2.5	9.9	17.8

Table 10: Google: Number of backlinks (domain pop)

SERP	1	2	3	∅	∅	$n-2$	$n-1$	n
US:								
Desktop	1035	326	185	**515**	**32**	38	18	40
Mobile	809	96	59	**321**	**178**	265	247	23
DE:								
Desktop	338	47	36	**140**	**69**	0	1	205
Mobile	180	46	11	**79**	**209**	0	382	245
IN:								
Desktop	925	160	158	**414**	**47**	63	21	58
Mobile	238	54	39	**110**	**81**	0	3	241
CF:								
Desktop	3817	256	4230	**2768**	**728**	911	936	336
Mobile	5112	397	0	**1836**	**388**	191	405	568

Table 11: Bing: Search contains title of webpage [data in pct.]

SERP	1	2	3	∅	∅	$n-2$	$n-1$	n
US:								
Desktop	100	100	94	**98**	**98**	93	100	100
Mobile	100	100	88	**96**	**97**	90	100	100
DE:								
Desktop	96	90	90	**92**	**84**	76	84	91
Mobile	96	90	90	**92**	**95**	90	94	100
IN:								
Desktop	95	80	63	**79**	**39**	30	34	53
Mobile	95	80	73	**83**	**74**	65	70	87

Table 12: Bing: Average number of words [data in thousands]

SERP	1	2	3	∅	∅	$n-2$	$n-1$	n
US:								
Desktop	2.4	1.8	1.6	**1.9**	**1.2**	1.3	1.2	1.1
Mobile	2.5	1.4	1.6	**1.8**	**1.3**	1.3	1.2	1.3
DE:								
Desktop	2.4	1.9	1.7	**2.0**	**1.0**	1.0	1.1	0.9
Mobile	2.4	1.5	1.3	**1.7**	**1.2**	1.0	1.2	1.3
IN:								
Desktop	1.9	1.9	1.3	**1.7**	**0.9**	1.0	0.9	0.8
Mobile	1.7	1.7	1.3	**1.6**	**1.5**	1.8	1.5	1.2

4.3.2 Bing

This section analyzes the results of the Bing search.

Search contains title of webpage: Similar to the results for Google, for Bing in general in the industrialized countries Germany and the United States, search queries are more likely to appear in the title of the webpage in contrast to the case in India (see Table 11). Nevertheless, it is noticeable that the difference between the best ranked search results and the worst ranked search results is not very big. The only exception is desktop search in India, where the best-ranked results almost twice as often carry a title, which is included in the search query, as opposed to the last 30 search results.

Average number of words: In contrast to the previous analysis of the average number of words on Google a much clearer picture is emerging from Table 12 for Bing. Accordingly, a word count of around 1800 words (as determined by Dean for Google in 2016 [6]) on average is a positive property when it comes to being well ranked by Bing.

Table 13: Bing: Webpage is encrypted [data in pct.]

SERP	1	2	3	∅	∅	$n-2$	$n-1$	n
US:								
Desktop	51	34	42	**42**	**43**	39	40	49
Mobile	48	36	43	**42**	**30**	41	25	25
DE:								
Desktop	46	37	42	**42**	**49**	52	43	52
Mobile	45	41	38	**41**	**45**	39	46	49
IN:								
Desktop	30	20	38	**29**	**41**	41	46	37
Mobile	37	46	41	**41**	**34**	36	30	35

Table 14: Bing: Webpage is responsive [data in pct.]

SERP	1	2	3	∅	∅	$n-2$	$n-1$	n
US:								
Desktop	96	91	94	**94**	**87**	84	91	85
Mobile	90	94	90	**91**	**84**	85	85	81
DE:								
Desktop	96	91	94	**94**	**82**	81	82	83
Mobile	94	91	85	**90**	**86**	84	88	87
IN:								
Desktop	94	94	90	**93**	**78**	79	84	72
Mobile	94	86	89	**90**	**89**	94	85	89

Webpage is encrypted: Unlike Google, HTTPS encrypted webpages does not seem to be a special property when it comes to tweaking a webpage to obtain a better ranking by Bing (see Table 13).

Webpage is responsive: In Table 14, it is noticeable that the top ranked search results are always slightly ahead for both desktop and mobile search. However, the difference is not really big. A slightly more significant picture emerges when the top 10 search results are compared to the last 10 results. Again, the same tendency applies, only a little bit more clear.

Webpage is an AMP document: The data in Table 15 indicates that Bing likes to prefer AMP documents neither for desktop nor for mobile searches. However, as AMP documents are very rarely among the search results in mobile search in India, Bing does not prefer to index AMP documents for mobile search, but lists them if there is no standard HTML version of the webpage.

Average URL length: According to the data in Table 16, a shorter URL is desirable if a webpage should rank better. Only India seems to be a bit out of line for the desktop search.

Table 15: Bing: Webpage is an AMP document [data in pct.]

SERP	1	2	3	∅	∅	$n-2$	$n-1$	n
US:								
Desktop	0	0	0	**0**	**0**	0	0	0
Mobile	0	0	0	**0**	**0**	0	0	0
DE:								
Desktop	0	0	0	**0**	**0**	0	0	0
Mobile	0	0	0	**0**	**0**	0	0	0
IN:								
Desktop	0	0	0	**0**	**0**	0	0	0
Mobile	0	0	0	**0**	**1**	1	1	0

Table 16: Bing: Average URL length

SERP	1	2	3	∅	∅	$n-2$	$n-1$	n
US:								
Desktop	43	52	51	**49**	**59**	60	58	58
Mobile	46	52	51	**50**	**59**	57	63	58
DE:								
Desktop	53	66	65	**61**	**61**	61	62	60
Mobile	54	62	69	**62**	**66**	63	67	68
IN:								
Desktop	53	60	63	**59**	**50**	53	50	48
Mobile	49	61	67	**59**	**70**	70	72	67

Table 17: Bing: Average page load time [data in seconds]

SERP	1	2	3	∅	∅	$n-2$	$n-1$	n
US:								
Desktop	4.7	5.3	5.4	**5.1**	**5.2**	5.8	5.2	4.6
Mobile	3.6	4.6	4.8	**4.3**	**5.0**	5.2	4.7	5.1
DE:								
Desktop	4.3	4.7	4.8	**4.6**	**5.3**	5.0	5.5	5.4
Mobile	3.8	4.3	3.6	**3.9**	**3.6**	3.7	3.9	3.2
IN:								
Desktop	4.5	5.6	5.7	**5.3**	**4.4**	3.7	4.1	5.4
Mobile	4.7	5.6	5.7	**5.3**	**5.2**	5.7	5.2	4.6

Average page load time: We recognized in our experiments, as well as previously for Google, that the worse ranked webpages do not always load slower than higher ranked pages by Bing (see Table 17). Like for Google before, the results of the mobile search are much faster for the best ranked as well as the worst ranked results than the results of the desktop search. However, an exception is India, where, on average, the search results of the desktop search load faster.

Number of backlinks: Table 18 contains the number of backlinks in the form of the link pop. Overall, a good majority of the better ranked webpages can be seen. Only the mobile search in India seems to be the exception. However, the results of the analysis are completely modified when looking at Table 19 and the backlinks in the form of the domain pop.

Table 18: Bing: Number of backlinks (link pop) [data in thousands]

SERP	1	2	3	∅	∅	$n-2$	$n-1$	n
US:								
Desktop	38.1	6.0	6.2	**16.8**	**4.2**	1.1	8.1	3.5
Mobile	15.2	6.4	2.5	**8.0**	**3.0**	2.5	6.1	0.3
DE:								
Desktop	3.1	4.4	0.7	**2.7**	**0.8**	0.0	0.2	2.2
Mobile	2.8	0.8	0.3	**1.3**	**0.1**	0.1	0.1	0.2
IN:								
Desktop	7.6	2.7	15.6	**8.6**	**2.4**	0.1	0.3	6.7
Mobile	2.6	1.2	0.2	**1.3**	**5.2**	0.0	14.4	1.1

Table 19: Bing: Number of backlinks (domain pop)

SERP	1	2	3	∅	∅	$n-2$	$n-1$	n
US:								
Desktop	1371	187	241	**600**	**202**	112	342	152
Mobile	576	186	147	**303**	**157**	151	284	35
DE:								
Desktop	289	208	64	**187**	**52**	10	33	112
Mobile	255	49	43	**116**	**23**	22	30	18
IN:								
Desktop	501	211	689	**467**	**111**	21	20	291
Mobile	230	69	68	**122**	**79**	14	206	18

4.3.3 Yahoo

We describe the analysis results of the Yahoo search in this section.

Search contains title of webpage: Like for the other search engines, in general in the industrialized countries, Germany and the United States, search queries rather appear in the title of the webpage in contrast to the case in India (see Table 20). Nevertheless, it is again noticeable that the difference between the best ranked search results and the worst ranked search results is not very big.

Average number of words: Similar to the results for Bing, all in all, a word count of around 1800 words can be seen as a positive property when it comes to being well ranked by Yahoo (see Table 21).

Table 20: Yahoo: Search contains title of webpage [data in pct.]

SERP	1	2	3	∅	∅	$n-2$	$n-1$	n
US:								
Desktop	100	100	92	**97**	**99**	97	100	100
Mobile	100	100	90	**97**	**95**	85	100	100
DE:								
Desktop	95	90	90	**92**	**91**	88	90	96
Mobile	95	90	90	**92**	**80**	76	80	85
IN:								
Desktop	94	80	65	**80**	**74**	63	75	85
Mobile	96	80	65	**80**	**72**	57	75	85

Table 21: Yahoo: Average number of words [data in thousands]

SERP	1	2	3	∅	∅	$n-2$	$n-1$	n
US:								
Desktop	2.6	1.9	1.6	**2.0**	**1.5**	2.0	1.2	1.3
Mobile	2.5	1.4	1.4	**1.8**	**1.2**	1.6	0.8	1.1
DE:								
Desktop	2.5	1.9	1.7	**2.0**	**1.2**	1.3	1.2	1.1
Mobile	2.2	1.7	1.4	**1.7**	**1.2**	1.2	1.1	1.4
IN:								
Desktop	1.9	2.0	1.3	**1.7**	**1.2**	1.1	1.3	1.2
Mobile	1.9	1.2	1.4	**1.5**	**1.3**	1.5	1.1	1.2

Table 22: Yahoo: Webpage is encrypted [data in pct.]

SERP	1	2	3	∅	∅	$n-2$	$n-1$	n
US:								
Desktop	51	38	48	**46**	**34**	34	34	33
Mobile	47	41	38	**42**	**40**	38	32	51
DE:								
Desktop	48	41	41	**43**	**48**	51	50	43
Mobile	45	44	37	**42**	**56**	46	57	66
IN:								
Desktop	34	29	35	**33**	**43**	38	49	41
Mobile	40	27	36	**34**	**42**	46	38	43

Webpage is encrypted: Similar to the results for Bing, HTTPS does not seem to be a special property when it comes to tweaking a webpage to obtain a better ranking by Yahoo (see Table 22).

Webpage is responsive: According to the data in Table 23 both, the desktop and mobile search, the top ranked search results are always slightly ahead in terms of responsive web design. The difference is very similar to the previous analysis of the same property for Bing, but overall it is a little bit more striking.

Table 23: Yahoo: Webpage is responsive [data in pct.]

SERP	1	2	3	∅	∅	$n-2$	$n-1$	n
US:								
Desktop	96	91	92	**93**	**84**	81	85	85
Mobile	88	91	80	**86**	**82**	84	82	81
DE:								
Desktop	97	93	91	**94**	**85**	91	84	80
Mobile	87	91	89	**89**	**83**	82	87	79
IN:								
Desktop	92	97	92	**94**	**88**	88	89	86
Mobile	93	88	84	**88**	**82**	84	85	78

Table 24: Yahoo: Webpage is an AMP document [data in pct.]

SERP	1	2	3	∅	∅	$n-2$	$n-1$	n
US:								
Desktop	0	0	0	**0**	**0**	0	0	0
Mobile	0	0	0	**0**	**0**	0	0	0
DE:								
Desktop	0	0	0	**0**	**0**	0	0	0
Mobile	0	0	0	**0**	**0**	0	0	0
IN:								
Desktop	0	0	0	**0**	**0**	0	0	0
Mobile	0	0	0	**0**	**0**	0	0	0

Table 25: Yahoo: Average URL length

SERP	1	2	3	∅	∅	$n-2$	$n-1$	n
US:								
Desktop	46	50	53	**50**	**57**	54	57	59
Mobile	47	50	52	**50**	**66**	57	59	81
DE:								
Desktop	53	65	65	**61**	**67**	68	66	68
Mobile	128	63	67	**86**	**84**	64	64	124
IN:								
Desktop	49	61	64	**58**	**67**	67	68	65
Mobile	82	57	70	**70**	**75**	62	68	96

Webpage is an AMP document: Yahoo does not list any AMP documents at least for the keywords and countries considered in this paper (see Table 24).

Average URL length: According to the experimental data in Table 25, even with Yahoo a shorter URL is desirable. Only the mobile search in Germany is slightly out of line.

Average page load time: The worse ranked webpages do not always load slower than higher ranked webpages (see Table 26). This is also the case with the other two search engines. Also, as previously stated, the results of

Table 26: Yahoo: Average page load time [data in seconds]

SERP	1	2	3	∅	∅	$n-2$	$n-1$	n
US:								
Desktop	5.2	5.2	6.2	**5.6**	**5.6**	5.9	5.9	5.0
Mobile	4.0	4.3	5.0	**4.4**	**4.2**	4.6	4.0	3.9
DE:								
Desktop	4.5	5.0	4.6	**4.7**	**4.7**	5.2	4.6	4.4
Mobile	2.7	5.1	4.1	**4.0**	**4.1**	4.8	4.5	3.2
IN:								
Desktop	5.0	6.6	6.4	**6.0**	**5.3**	5.3	5.0	5.5
Mobile	3.4	4.3	4.5	**4.0**	**4.4**	4.7	4.3	4.1

Table 27: Yahoo: Number of backlinks (link pop) [data in thousands]

SERP	1	2	3	∅	∅	$n-2$	$n-1$	n
US:								
Desktop	26.7	14.2	6.2	**15.7**	**3.7**	1.0	1.6	8.5
Mobile	14.5	12.2	2.8	**9.8**	**0.9**	1.2	0.4	1.1
DE:								
Desktop	3.1	4.5	0.3	**2.6**	**2.2**	3.7	0.1	2.7
Mobile	2.8	0.8	0.3	**1.3**	**0.2**	0.2	0.1	0.3
IN:								
Desktop	7.7	3.8	14.4	**8.7**	**1.8**	0.2	0.1	5.1
Mobile	2.6	1.0	0.6	**1.4**	**5.7**	0.2	0.4	16.4

the mobile search for both the best ranked and the worst ranked results are always faster than the results of the desktop search.

Number of backlinks: Table 27 contains the number of backlinks in the form of the link pop. Overall, a good majority of the better ranked webpages can be seen. Like Bing before, only the mobile search in India seems to be an exception. However, the analysis results are completely changed when looking at Table 28, i.e. the domain pop.

4.3.4 All Countries

In this section, the results of all countries and search engines are looked at. In order to be able to correctly compare the overall results of Google with those of Bing and Yahoo, the Central African Republic is not considered since data for the Central African Republic could not be determined using Bing and Yahoo.

Search contains title of webpage: According to Table 29 we retrieve for all three search engines similar results: About 90% of the top 30 search results have one, at least partly, match between the title of the webpage

Table 28: Yahoo: Number of backlinks (domain pop)

SERP	1	2	3	∅	∅	$n-2$	$n-1$	n
US:								
Desktop	1024	509	163	**565**	**175**	64	90	371
Mobile	566	346	173	**362**	**91**	114	51	108
DE:								
Desktop	288	206	41	**178**	**168**	340	22	142
Mobile	255	45	48	**116**	**43**	76	17	37
IN:								
Desktop	530	268	615	**471**	**59**	29	23	126
Mobile	234	69	93	**132**	**107**	25	36	259

Table 29: Overall: Search contains title of webpage [data in pct.]

SERP	1	2	3	∅	∅	$n-2$	$n-1$	n
Google:								
Desktop	98	90	82	**90**	**91**	86	91	97
Mobile	97	90	82	**90**	**89**	81	91	94
Bing:								
Desktop	97	90	82	**90**	**73**	66	73	81
Mobile	97	90	84	**90**	**89**	82	88	96
Yahoo:								
Desktop	96	90	82	**89**	**88**	83	88	94
Mobile	97	90	82	**90**	**83**	73	85	90

and the searched keywords. Google remains relatively constant even when the results are lower ranked, while there is a slight decline for Bing and Yahoo.

Average number of words: According to the data in Table 30 across the analyzed search engines, more words indicate a better ranking. Across these search engines it can also be seen that the results of the mobile search on average have fewer words than the results of the desktop search. However, this only applies to the top 30 search results. For the last 30 search results, this is not always the case.

Webpage is encrypted: The use of encryption with HTTPS seems to play a bigger role only for Google (see Table 31). The other search engines do not seem to pay much attention to the use of HTTPS.

Webpage is responsive: The usage of responsive web design, as seen in Table 32 in the overall consideration, seems to be an advantage for obtaining a good ranking in all examined search engines. However, overall it can also be concluded that the proportion of responsive designed websites for the results of the mobile search is smaller than for the results of the desktop search.

Table 30: Overall: Average number of words [data in thousands]

SERP	1	2	3	∅	∅	$n-2$	$n-1$	n
Google:								
Desktop	2.1	1.6	1.3	**1.7**	**1.4**	1.5	1.6	1.1
Mobile	2.1	1.2	1.1	**1.4**	**1.3**	1.1	1.3	1.4
Bing:								
Desktop	2.2	1.9	1.5	**1.9**	**1.0**	1.1	1.0	0.9
Mobile	2.2	1.5	1.4	**1.7**	**1.3**	1.3	1.3	1.3
Yahoo:								
Desktop	2.3	1.9	1.5	**1.9**	**1.3**	1.4	1.2	1.2
Mobile	2.2	1.5	1.4	**1.7**	**1.2**	1.4	1.0	1.2

Table 31: Overall: Webpage is encrypted [data in pct.]

SERP	1	2	3	∅	∅	$n-2$	$n-1$	n
Google:								
Desktop	53	41	41	**45**	**41**	42	46	34
Mobile	51	42	46	**46**	**38**	39	40	35
Bing:								
Desktop	42	30	41	**38**	**44**	44	43	46
Mobile	43	41	41	**42**	**36**	39	34	36
Yahoo:								
Desktop	44	36	41	**40**	**41**	41	44	39
Mobile	44	37	37	**39**	**46**	43	42	53

Table 32: Overall: Webpage is responsive [data in pct.]

SERP	1	2	3	∅	∅	$n-2$	$n-1$	n
Google:								
Desktop	93	93	88	**91**	**83**	85	86	79
Mobile	87	89	91	**89**	**79**	84	80	72
Bing:								
Desktop	95	92	93	**93**	**82**	81	86	80
Mobile	93	90	88	**90**	**87**	88	86	86
Yahoo:								
Desktop	95	94	92	**94**	**86**	87	86	84
Mobile	89	90	84	**88**	**82**	83	85	79

Webpage is an AMP document: Table 33 contains data about the use of AMP documents. Like in the individual analyses, it can be seen that AMP documents currently play only for Google a significant role.

Average URL length: Overall, the average length of the URLs as given in Table 34 draws a very clear picture, which applies equally to all search engines considered in this work: A shorter URL is quite conducive to achieve a better ranking on search engines.

Table 33: Overall: Webpage is an AMP document [data in pct.]

SERP	1	2	3	∅	∅	$n-2$	$n-1$	n
Google:								
Desktop	0	0	0	**0**	**0**	0	0	0
Mobile	24	25	19	**23**	**4**	6	4	2
Bing:								
Desktop	0	0	0	**0**	**0**	0	0	0
Mobile	0	0	0	**0**	**0**	0	0	0
Yahoo:								
Desktop	0	0	0	**0**	**0**	0	0	0
Mobile	0	0	0	**0**	**0**	0	0	0

Table 34: Overall: Average URL length

SERP	1	2	3	∅	∅	$n-2$	$n-1$	n
Google:								
Desktop	65	67	65	**66**	**82**	81	81	83
Mobile	65	70	68	**68**	**79**	81	79	78
Bing:								
Desktop	50	59	60	**56**	**57**	58	57	55
Mobile	50	58	62	**57**	**65**	63	67	64
Yahoo:								
Desktop	49	59	61	**56**	**64**	63	64	64
Mobile	86	57	63	**69**	**75**	61	64	100

Table 35: Overall: Average page load time [data in seconds]

SERP	1	2	3	∅	∅	$n-2$	$n-1$	n
Google:								
Desktop	4.7	5.9	5.9	**5.5**	**4.6**	4.7	4.7	4.6
Mobile	3.1	3.6	3.7	**3.5**	**3.9**	4.1	3.8	3.9
Bing:								
Desktop	4.5	5.2	5.3	**5.0**	**5.0**	4.8	4.9	5.1
Mobile	4.0	4.8	4.7	**4.5**	**4.6**	4.9	4.6	4.3
Yahoo:								
Desktop	4.9	5.6	5.7	**5.4**	**5.2**	5.4	5.2	5.0
Mobile	3.4	4.5	4.5	**4.1**	**4.2**	4.7	4.2	3.7

Average page load time: According to the data in Table 35 and like previously seen in the individual analyses, lower ranked webpages do not always load slower than higher ranked webpages. Also, as previously stated, for both the best ranked and the worst ranked results, the mobile search is always faster in terms of load time than the desktop search results. Overall the mobile search results for the top 30 results have a faster load time than the worst 30 search results. However, the difference is usually very small.

Table 36: Overall: Number of backlinks (link pop) [data in thousands]

SERP	1	2	3	∅	∅	$n-2$	$n-1$	n
Google:								
Desktop	15.2	4.0	3.2	**7.5**	**1.8**	0.1	0.1	5.3
Mobile	9.1	1.8	0.3	**3.7**	**21.4**	1.1	57.5	5.7
Bing:								
Desktop	16.2	4.4	7.5	**9.4**	**2.5**	0.4	2.9	4.1
Mobile	6.9	2.8	1.0	**3.6**	**2.8**	0.9	6.9	0.5
Yahoo:								
Desktop	12.5	7.5	7.0	**9.0**	**2.6**	1.6	0.6	5.4
Mobile	6.6	4.6	1.2	**4.2**	**2.3**	0.5	0.3	5.9

Table 37: Overall: Number of backlinks (domain pop)

SERP	1	2	3	∅	∅	$n-2$	$n-1$	n
Google:								
Desktop	766	178	126	**357**	**49**	34	13	101
Mobile	409	65	36	**170**	**156**	88	211	170
Bing:								
Desktop	720	202	331	**418**	**122**	48	132	185
Mobile	354	101	86	**180**	**86**	62	173	24
Yahoo:								
Desktop	614	328	273	**405**	**134**	144	45	213
Mobile	352	153	105	**203**	**81**	72	35	135

Number of backlinks: Like in the previous single analyses, only the mobile search on Google is an outlier (see Table 36), which cannot be acknowledged when looking at Table 37 in terms of the domain pop.

4.3.5 Discussion of Analysis Results

In this section, the results from the previous sections are summarized for the desktop search and the mobile search. Here, the examined webpage properties are evaluated in terms of their influence on the ranking of a webpage. To do this evaluation, the following rating scale is used:

- 0 Property has no noticeable impact on the ranking. The difference between the average of the top 30 results and the 30 worst-ranked results is below 5%
- \+ Property has little impact on the ranking. The difference between the average of the top 30 results and the 30 worst-ranked results is between 5% and 15%
- ++ Property has significant impact on the ranking. The difference between the average of the top 30 results and the 30 worst-ranked results is over 15%

Overview of the desktop search: In Table 38, the analysis of desktop search results is summarized and illustrated. It is striking that the load time plays almost no role. It is also noticeable that AMP has absolutely no influence on the ranking of results for the desktop search. It seems amazing how much the length of the URL influences the rankings across all search engines. It is also surprising that the occurrence of the page title in the search term does not seem to matter. Only Bing seems to put more emphasis on this webpage property. HTTPS encrypted webpages seem to lead to a slightly better ranking only on Google. Responsive Web Design (RWD), on the other hand, is very important to all search engines, which might not have been expected for the desktop search. A larger number of words also seems to be a positive aspect to all search engines considered in this paper. The number of backlinks is very important to all search engines.

Overview of the mobile search: Table 39 summarizes the analysis of mobile search results. In contrast to the desktop search, the load time seems to have medium to significant influence on the webpage ranking. In particular, Google seems to put special emphasis on fast webpages when it comes to mobile search. This does not seem to be surprising, as Google is driving AMP forward as a technology, which focuses on fast loading webpages. For the other search engines no ranking influence of AMP is detected. The use of HTTPS for mobile search seems to have medium influence only for Google and Bing. Yahoo does not seem to care about this webpage property. Responsive web design seems to be important to all search engines. However, it is noticeable that the influence of this webpage property on the mobile search seems to be slightly lower than the impact on the desktop search, which is surprising since responsive web design is especially important on mobile devices. The length of the URL seems as important to the mobile search as it is to the desktop search. The occurrence of the webpage title in the search term also draws a similar picture. It seems like that in mobile search this property is also not important. Further parallels to the desktop search is found in the number of words and the backlinks, too. Both are also seen as a webpage property with significant impact on mobile search.

5 Summary and Conclusions

In Section 1 an introduction to search engine optimization is provided and further explored in Sections 2 and 3. In section 4 we discuss the selection of search engines, countries, search terms and webpage properties for our analysis. Afterwards we analyze in detail the data determined by our tool developed within the scope of

Table 38: Summary of the desktop search

Criteria	Load time	HTTPS	RWD	AMP	Backlinks	URL	Title	Words
Google	0	+	+	0	++	++	0	++
US	0	++	+	0	++	++	0	0
DE	0	0	0	0	+	++	0	++
IN	0	++	++	0	++	++	0	++
Bing	0	0	+	0	++	0	++	++
US	0	0	+	0	++	++	0	++
DE	+	0	+	0	++	0	+	++
IN	0	0	++	0	++	0	++	++
Yahoo	0	0	+	0	++	+	0	++
US	0	++	+	0	++	+	0	++
DE	0	0	+	0	+	+	0	++
IN	0	0	+	0	++	+	+	++

Table 39: Summary of the mobile search

Criteria	Load time	HTTPS	RWD	AMP	Backlinks	URL	Title	Words
Google	+	++	+	++	+	+	0	+
US	++	++	0	++	++	++	0	0
DE	0	0	+	++	0	+	0	++
IN	++	++	++	++	+	+	+	++
Bing	0	+	0	0	++	+	0	++
US	+	++	+	0	++	+	0	++
DE	0	0	+	0	++	+	0	++
IN	0	++	0	0	+	++	+	+
Yahoo	0	0	+	0	++	+	+	++
US	0	+	+	0	++	++	0	++
DE	0	0	+	0	++	0	+	++
IN	+	0	+	0	+	+	+	++

this paper. The results of this analysis are sometimes not quite as expected. As an example it should be noted again that for all search engines the proportion of webpages that are responsive designed are always slightly lower for the mobile search than for the desktop search. The difference is never really big, but it seems more likely that the share for the mobile search will be bigger, as responsive web design is particularly important for mobile devices. We further recognize in our analysis that encrypted webpages seem to have only a slight advantage affecting the ranking on Google. On Bing and Yahoo, no real difference is found between best-ranked and lowest-ranked webpages referring to encryption. AMP documents are currently only used by Google. There, however, it seems to be an advantage especially for mobile search. For the desktop search on the other hand AMP documents are also ignored by Google and thus do not result in better rankings. In conclusion, there are no significant surprises regarding the other examined webpage properties. Furthermore, within the scope of our analysis, there are no noticeable country-specific differences referring to webpage quality.

In future work, we plan to extend our analysis to more properties of webpages, other search engines including metasearch engines and specialize the experiments to target webpages of different areas like entertainment, news, webshops and research. We also want to investigate changes in website rankings over time especially after updates of the ranking algorithms. We may also extend our research to other types of search engines like searching for scientific articles via Google Scholar[12], PubMed[13] and DBLP[14] and searching for images and videos.

REFERENCES

[1] J. Bar-Ilan, M. Mat-Hassan, and M. Levene, "Methods for comparing rankings of search engine results," May 2005. [Online]. Available: http://cds.cern.ch/record/835959

[2] J. Bar-Ilan, "Comparing rankings of search results on the web," *Information Processing & Management*, vol. 41, no. 6, pp. 1511 – 1519, 2005, special Issue on Infometrics.

[3] K. Bharat and G. A. Mihaila, "Hilltop: A Search Engine based on Expert Documents," http://ftp.cs.toronto.edu/pub/reports/csrg/405/hilltop.html, 2000.

[4] A. Bifet, C. Castillo, P.-A. Chirita, and I. Weber, "An analysis of factors used in search engine ranking," in *Proceedings of the Workshop on Adversarial IR on the Web*, 2005.

[5] B. Dean, "Google's 200 ranking factors: The complete list," https://backlinko.com/google-ranking-factors, Nov. 05 2016.

[6] B. Dean, "We analyzed 1 million google search results. here's what we learned about SEO," http:

[12] https://scholar.google.com/

[13] https://www.ncbi.nlm.nih.gov/pubmed/

[14] https://dblp.uni-trier.de/

//backlinko.com/search-engine-ranking, Sep. 02 2016.

[7] G. Egri and C. Bayrak, "The role of search engine optimization on keeping the user on the site," *Procedia Computer Science*, vol. 36, pp. 335 – 342, 2014.

[8] M. P. Evans, "Analysing google rankings through search engine optimization data," *Internet Research*, vol. 17, no. 1, pp. 21–37, 2007.

[9] A. Gandour and A. Regolini, "Web site search engine optimization: a case study of fragfornet," *Library Hi Tech News*, vol. 28, no. 6, pp. 6–13, 2011.

[10] Google Inc., "Google does not use the keywords meta tag in web ranking," https://webmasters.googleblog.com/2009/09/google-does-not-use-keywords-meta-tag.html, 2009.

[11] Google Inc., "Search Engine Optimization Starter Guide," https://static.googleusercontent.com/media/www.google.com/de//webmasters/docs/search-engine-optimization-starter-guide.pdf, 2010.

[12] Google Inc., "Official Google Webmaster Central Blog: Using page speed in mobile search ranking," https://webmasters.googleblog.com/2018/01/using-page-speed-in-mobile-search.html, 2018.

[13] V. N. Gudivada, D. Rao, and J. Paris, "Understanding Search Engine Optimization," *IEEE Computer*, vol. 48, no. 10, pp. 43–52, 2015.

[14] International Monetary Fund, "World Economic Outlook Database — WEO Groups and Aggregates Information," http://www.imf.org/external/pubs/ft/weo/2010/02/weodata/groups.htm, 2010.

[15] International Monetary Fund, "World Economic Outlook Database — GDP per capita, current prices," http://www.imf.org/external/datamapper/PPPPC@WEO/THA, 2017.

[16] Internet Live Stats, "Total number of Websites," http://www.internetlivestats.com/total-number-of-websites/, accessed 30th May 2018.

[17] M. Kendall and J. Gibbons, *Rank correlation methods*, ser. A Charles Griffin Book. E. Arnold, 1990.

[18] T. Mavridis and A. L. Symeonidis, "Identifying valid search engine ranking factors in a web 2.0 and web 3.0 context for building efficient seo mechanisms," *Eng. Appl. Artif. Intell.*, vol. 41, no. C, pp. 75–91, May 2015.

[19] P. J. Meyers, "What Keywords Do I Rank For?" https://moz.com/blog/what-keywords-do-i-rank-for, Mar. 09 2011.

[20] P. J. Meyers, "Face-off - 4 Ways to De-personalize Google," https://moz.com/blog/face-off-4-ways-to-de-personalize-google, Jan. 25 2012.

[21] Moz Inc., "Google Algorithm Change History," https://moz.com/google-algorithm-change, accessed 30th May 2018.

[22] N. Spirin and J. Han, "Survey on Web Spam Detection: Principles and Algorithms," *ACM SIGKDD Exploration Newsletter,*, vol. 13, no. 2, pp. 50–64, 2012.

[23] U. N. D. of Economic and P. D. Social Affairs, "World population prospects," https://esa.un.org/unpd/wpp/DataQuery/, accessed 29th May 2018.

[24] L. Page, "Method for node ranking in a linked database," Sep. 4 2001, US Patent 6,285,999. [Online]. Available: https://www.google.com/patents/US6285999

[25] S. P. Patil, B. Pawar, and A. S. Patil, "Search Engine Optimization: A Study," *Research Journal of Computer and Information Technology Sciences*, vol. 1, no. 1, pp. 10–13, Feb. 2013.

[26] C. Ratcliff, "How to completely dominate Google's first page," https://searchenginewatch.com/sew/how-to/2438200/how-to-completely-dominate-googles-first-page, Dec. 09 2015.

[27] M. Southern, "Bing's Top Searches of 2016," https://www.searchenginejournal.com/bings-top-searches-2016/181748/, Dec. 19 2016.

[28] Statista Inc., "Market share of the leading search engines in Germany from 2014 to 2016," https://www.statista.com/statistics/445002/market-shares-leading-search-engines-germany/, Mar. 2016.

[29] Statista Inc., "Desktop and mobile search market share of search engines in Germany in August 2017," https://www.statista.com/statistics/445974/search-engines-market-share-of-desktop-and-mobile-search-germany/, Sep. 2017.

[30] A. J. Su, Y. C. Hu, A. Kuzmanovic, and C. K. Koh, "How to improve your google ranking: Myths and reality," in *2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, vol. 1, Aug 2010, pp. 50–57.

[31] United Nations Development Programme, "Human Development Report 2016: Human Development for Everyone," http://hdr.undp.org/sites/default/files/2016_human_development_report.pdf, 2016.

[32] A. Wicks, "Yahoo Reveals 2016 Year in Review: Top News Stories, Pop Culture, Sports Moments, Fashion Trends, and More," https://yahoo.tumblr.com/post/154119318524/yahoo-reveals-2016-year-in-review-top-news, Dec. 06 2016.

[33] J. Zhang and A. Dimitroff, "The impact of metadata implementation on webpage visibility in search engine results (part ii)," *Inf. Process. Manage.*, vol. 41, no. 3, pp. 691–715, May 2005.

AUTHOR BIOGRAPHIES

Sinan Babayigit studies Computer Science at the University of Lübeck. He wrote his Bachelor's thesis about the topic of search engine optimization and described the development of a web application for the analysis of properties of best ranked webpages. The tool was used to conduct the analysis described in this paper. His interests include everything around the web and the web technologies.

Sven Groppe earned his diploma degree in Informatik (Computer Science) in 2002 and his Doctor degree in 2005 from the University of Paderborn. He earned his habilitation degree in 2011 from the University of Lübeck. He worked in the European projects B2B-ECOM, MEMPHIS, ASG and TripCom. He was a member of the DAWG W3C Working Group, which developed SPARQL. He was the project leader of the DFG project LUPOSDATE, an open-source Semantic Web database, and one of the project leaders of two research projects, which research on FPGA acceleration of relational and Semantic Web databases. He is also the chair of the Semantic Big Data workshop series, which is affiliated with the ACM SIGMOD conference (so far 2016 to 2018), and of the Very Large Internet of Things workshop in conjunction with the VLDB conference in 2017 and 2018. His research interests include databases, Semantic Web, query and rule processing and optimization, Cloud Computing, peer-to-peer (P2P) networks, Internet of Things, data visualization and visual query languages.