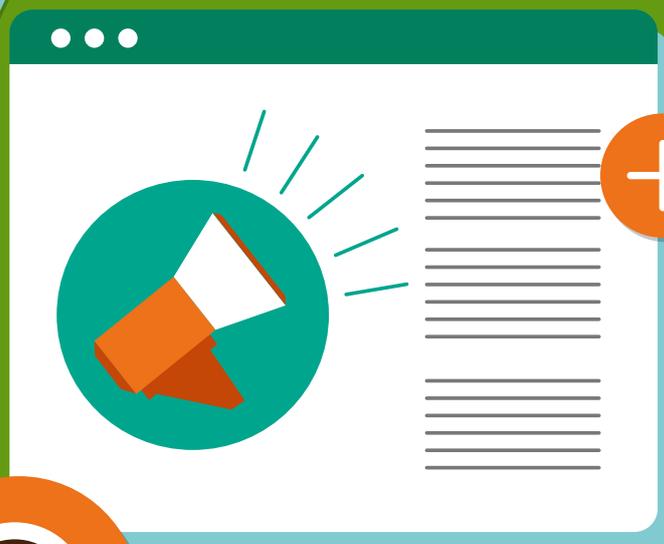


Whitepaper



SEO Ranking Factors and Rank Correlations 2014

- Google U.S. -

Authors

Marcus Tober
Dr. Leonhard Hennig
Daniel Furch



Whitepaper

SEO Ranking Factors and Rank Correlations 2014

- Google U.S. -



Authors: Marcus Tober, Dr. Leonhard Hennig, Daniel Furch

Note

This white paper deals with the definition and evaluation of factors that have a high rank correlation-coefficient with organic search results, and acts as a deeper analysis of search engine algorithms. Data collection, including evaluation, refers to ranking factors for the U.S. index of Google.com in 2014.

In addition, correlations and other factors, based on mean and median values and trends from previous years, with respect to their relevance for forward search result positions, have been interpreted.

About Searchmetrics and Searchmetrics Suite™

Searchmetrics is the pioneer and leading global enterprise platform for search experience optimization. Search Experience Optimization combines SEO, content, social media, PR and analysis to create the foundation for developing and executing an organic search strategy. It places the spotlight on the customer, contributing to a superior and memorable purchase experience.

Enterprises and agencies use the Searchmetrics Suite to plan, execute, measure and report on their digital marketing strategies. Supported by a continually updated global database, Searchmetrics answers the key questions asked by SEO professionals and digital marketers. It delivers a wealth of forecasts, analytic insights and recommendations that boost visibility and engagement, and increase online revenue. Many respected brands, such as T-Mobile, eBay, Siemens and Symantec, rely on the Searchmetrics Suite.

Searchmetrics has offices in Berlin, San Mateo, New York, London, and Paris, and is backed by Holzbrinck Digital, Neuhaus Partners and Iris Capital.

Infographic: Google Ranking Factors in 2014

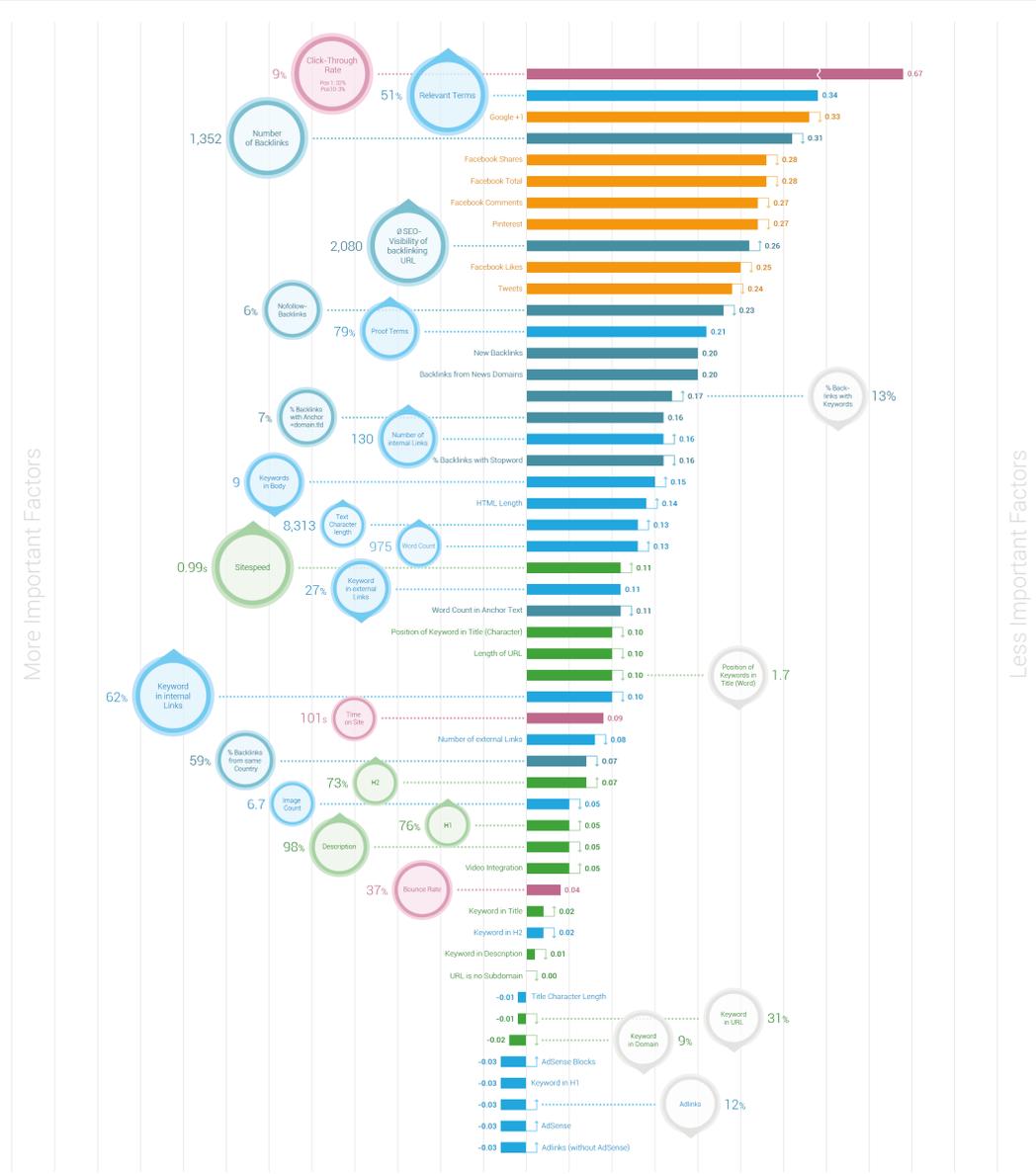
SEARCHMETRICS

RANKING-FACTORS 2014

Correlation of Top 30 / Average of Top 10

Google.com

HOW TO READ THE GRAPHIC



What do the higher-ranking sites in the top positions in Google have in common and what differentiates them from those ranking behind?

To investigate this question, we have analyzed the search results for 10,000 keywords according to the existence and extent of certain properties. Here are the results represented both as correlation (Spearman) calculated for the top 30 and as average values based on the top 10.

Download the whole study at searchmetrics.com/us2014



Summary of Results: Abstract

High quality, relevant content is increasingly the focus of search. This type of content ranks better on average, and is identifiable by properties such as a higher word-count and semantically comprehensive wording, as well as often being enriched by other media, such as images or video. Regarding content features analyzed in this study, both correlations and average values have increased compared to 2013.

For on-page SEO, the keyword is still vital, but the trend is moving away from a single subject/keyword towards entire topics. Almost all keyword factors diminish slightly regarding averages with this approach, but correlations are not always affected negatively.

Fundamental to a good ranking is the technical performance of a page. These factors include robust site architecture with good internal links, short loading times and the presence of all relevant Meta tags, such as Title and Description

The correlation for technological features is often low, or near zero, but as these are often binary features with very high average values over the top thirty rankings, the performance of these factors to ensure good ranking is still mandatory.

In contrast, social signals stagnate easily, resulting in decreasing correlations for those networks investigated. Averages grew a little compared to 2013.

The quantity and quality of backlinks remains crucial. Here, many new features are introduced this year, and many features have been revised to improve the quality of the results. The correlation between ranking and backlink features is often high.

As in previous years, the importance of 'brand' runs through the study – but had to be remodeled and redefined this year because of new findings. Both small and large brands play a role – even if some of the respective brand properties vary greatly. Inclusion in Wikipedia remains highly significant.

For the first time this year user signals were also measured against search results. As expected, we found that correlation with „click-through rate“ is high, sites appearing in the Top 5 tend to have a lower „bounce rate,“ and a high „time-on-site“ is very likely to have a positive impact on rankings.

Please keep in mind, a good search position cannot simply be achieved by ticking a few of these boxes. The search engine assessment of the relevance of pages is always multifaceted, and ranking positions are based on the interaction of many different factors. Correlations described here are not to be interpreted as causal relationships across-the-board.

Content

Note	1
About Searchmetrics and Searchmetrics Suite™	3
Infographic: Google Ranking Factors in 2014	4
Summary of Results: Abstract	5
What Is A Ranking Factor?	9
Causality ≠ Correlation	10
Factors	10
The Data Used	11
Database For Searchmetrics Ranking Factors	11
Binary And Numerical Factors - Specification Versus Existence	13
Correlation Values Versus Average Values And Curves	13
Adapted Feature Calculations	13
Structure of the White Paper	13
At a glance: Infographic – Technology	15
1. ONPAGE TECHNOLOGY	16
Onpage Technology: Development Compared To Last Year	17
Still Few Keyword Domains In Search Results	18
Keyword Frequency In Title And Description	19
Content Digression: Keywords Are Rare In Hx-Headings	21
Position Of Keyword Quite Early In The Title	21
Page Load Times As An Important Performance Factor	22
Use Of Flash No Longer Widespread	24
Summary	24
At a glance: Infographic – Content	25
2. ONPAGE CONTENT	26
Onpage Content: development compared to last year	27
Readability Of Text - Flesch & Co.	31
2.2 Content length increases significantly (again)	32
HTML Length	33
Text Length	34
Number Of Sentences	35

2.3 Enrichment of content by media pays off	35
2.4 Internal Linking: Housekeeping is the beginning and the end	36
2.5 Adlinks: In total less advertising – except for Brands	38
Author Integrations In Google SERPs	40
Summary	41
At a glance: Infographic – Backlinks	42
3. BACKLINKS	43
Backlinks: Development Compared To Last Year	43
3.1 Link Source: The higher the quality, the better	44
The Number Of Backlinks Remains Important	44
Better Ranked URLs Have Much Better Links	45
Number Of News-Domain Links To The Homepage	47
Number Of Links From The Same Country	47
3.2 Link Quality: The need for quality continues to rise	48
New Backlinks	48
Share of Nofollow Links	48
Average Link Age	49
Homepage / Deep Link Distribution	49
3.3 Link Anchor – Link Diversity Becomes More Relevant	50
Keyword Anchor Texts Are Rare	50
Number Of Words In Anchor Text	52
Proportion Of Backlinks With Anchor Text = Domain.Tld	53
Proportion Of Brand Anchor Texts To Domain And URL	53
Summary	55
At a glance: Infographic – Social	56
4. SOCIAL SIGNALS	57
Social Signals: Developments In The Previous Year	57
Social Signal: Correlations Decreasing, But Still High	58
Summary	60
5. THE BRAND FACTOR AND WIKIPEDIA	61
Three Different Kinds Of Brands	61
Search Volume, Pagerank And Alexarank	66
Searchmetrics Page Strength (SPS) For The Homepage	68

„Brand Links“ - What Is A Brand And Are Anchor Texts A Brand-Reference?	69
Wikipedia: The Best-Optimized Universal Brand	70
The Top Positions Are Reserved For Brands - And Wikipedia	71
At a glance: Infographic – User Signals	72
6. USER AND TRAFFIC SIGNALS	73
7. DIGRESSION AND OUTLOOK: DESKTOP VS MOBILE	76
Summary of Searchmetrics SEO Ranking Factors 2014	79
Summary: The Main Developments For Ranking Factors In 2014	80

Definition, Influential Factors And Database

What Is A Ranking Factor?

Search engines work by using algorithms to evaluate websites by topic and relevance. This evaluation is used to structure pages in the search engine index, which ultimately results in user queries displaying the best possible ranking of the results displayed. The criteria for the evaluation of web pages and to produce this ranking are generally referred to as ranking factors.

The reasons for this are:

- | The exponentially rising number of documents on the Internet - and in the search index – makes it impossible to rank these pages without an automatic algorithm, despite the existence of human 'quality raters.'
- | This algorithm is both mandatory (order, after all, requires a pattern), and, at the same time, the best-kept secret in the Internet business, because for search engines, it is essential to keep the underlying factors that make up the algorithm strictly confidential. This inherent secrecy has less to do with competition between search engines than it has to do with more basic reasons: If the ways to obtain good rankings were widely known, they would become irrelevant as they would be constantly manipulated.

At the beginning of the search engine age, Google considered pages relevant for specific topics where the subject-associated search terms (keywords) were frequently used. Site operators soon took advantage of this knowledge and achieved very good positions in the SERPs¹ by 'stuffing' pages with keywords, enabling their often non-relevant pages to be found on well ranking positions for the intended search terms.

This generated not only real competition between search engines and SEOs, but produced the myth of the ranking factor. The goal of semantic search created a network of criteria that were initially strictly technical (e.g. the number of backlinks), but soon after less technical components were also added (e.g. user signals).

This development, along with the pursuit of the optimum result, has culminated in the constant evolution of ranking factors. The endless feedback loop of permanent-iterative update cycles is designed purely to generate search results that offer constant improvements to the individual searcher. The structure and complexity of ranking factors, added to the strong influence of user signals, is designed to produce the most relevant search experience for the user.

From a business perspective, long-term success can be achieved by using a sustainable business strategy based on incorporating relevant quality factors to maintain strong search positions. This approach means a disregard of negative influence options and a clear focus on relevant content, at the same time combating spam and short-termism.

¹ Search Engine Results Pages

Causality ≠ Correlation²

We are not Google. The analysis and evaluation of ranking factors using our data is not based on speculation, but upon well-founded interpretation of the facts; namely the evaluation and structuring of web site properties with high positions in the search results.

The ranking evaluations of Searchmetrics are exact, fresh and based on huge amounts of data. We aggregate billions of data points on a monthly basis, and provide an answer to the 64-million dollar question, – Which factors distinguish well-placed sites from those with lower positions in the Google search results?

We compare the properties of Web pages with the sum of their positions at Google and derive a structured list of more or less weighted factors. So, if in the front positions of the investigated SERPs, for example, many pages contain the keyword in the title tag, then we recognize it as a high correlation with a good ranking. These correlations can therefore offer conclusions as to which pages displayed in the top search results listings have features that are similar.

However, it is necessary to point out that correlations are not synonymous with causal relationships, and hence there is no guarantee that the respective factors actually have any impact on the ranking - or are ever used by Google as a signal. What we do at Searchmetrics, is to interpret these correlations.

The data speaks for itself. Unlike the definitive algorithmic ranking factors of Google and the size of their influence, the study data that we use is not secret. Thus, our analyses not only allow an assessment as to which factors are most likely to be involved in the ranking, but also how pronounced their influence may be. So we can move on from 'gut feeling' and begin to confirm or refute hypotheses by using the data in this study.

As mentioned, there are purely technical, or „hard,“ as well as less technical, or „soft“ ranking factors. To make this as clear as possible in our graphics, we grouped together these factors in our social, backlinks, onpage and content sections, even though there is slight overlap in some areas.

The term „ranking factor“ does sometimes lead to interpretations that are not always accurate. In our experience, we prefer to use the term “rank correlation coefficient” to describe the context of our analysis, as this is a much more accurate reflection of what it is we set out to prove.

Factors

The position and order of search results not only vary depending on the search terms entered (keywords), but are also influenced, among other things, by personal search history, which in turn is determined by user behavior itself. They also depend on factors such as whether the user is logged into Google, surfing Chrome, the user IP location, cookies with personal information stored on the computer, etc. Google itself varies over time, and to consider all of these factors would lead to an overly complex white paper, so this paper focuses on the variations of the types of data that Google itself uses, to give results that are unbiased and clear.

² See also: searchmetrics.com/en/what-is-a-ranking-factor/*

³ multiple times

The Data Used

The data used in this study was collected in 2014³ and then aggregated. Since this is a follow-up study to the studies carried out during 2012 and 2013, the changes in many factors from those of previous years are particularly interesting, because they argue strongly for the evolution of ranking factors. This evolution offers insights into trends for the future development, and relevance of the criteria chosen.

In addition, the range of criteria studied in 2014 was extended to allow for further analysis of existing factors, and to illuminate new areas of research. The focus this year was primarily on content factors, but backlink factors were also expanded for this year's report.

Database For Searchmetrics Ranking Factors

Our analysis is based on search results for a very large keyword set of 10,000 search terms for Google U.S. The pool is always the top 10,000 search terms by search volume, with strong focus on informational keywords in order not to distort the evaluations. In contrast to informational search terms, navigational keyword searches generally target one result only, for example, „Facebook login,“ making it the only one relevant to the searcher.

Our database for the Ranking Factor analyses are always the first three organic search results pages. The keyword sets from consecutive years coincide by more than 90 percent with the database from the previous year, as a rule. Here we have sought a middle ground, to take two factors into account, namely the preservation of the „greatest common denominator“ as the optimal basis for comparison with the previous study, and on the other hand, taking into account new keywords, which have grown in search volume in the top 10,000.

The database at Searchmetrics is always current. Therefore, new, relevant keywords are used for current analyses, such as „Samsung Galaxy S5“ or „iPhone 6,“, which did not exist previously.

The Most Important Correlation Coefficients At-A-Glance

As in our previous analysis, relationships between the various ranking factors are shown using a Spearman correlation. The most important correlations are shown in the following bar graph:

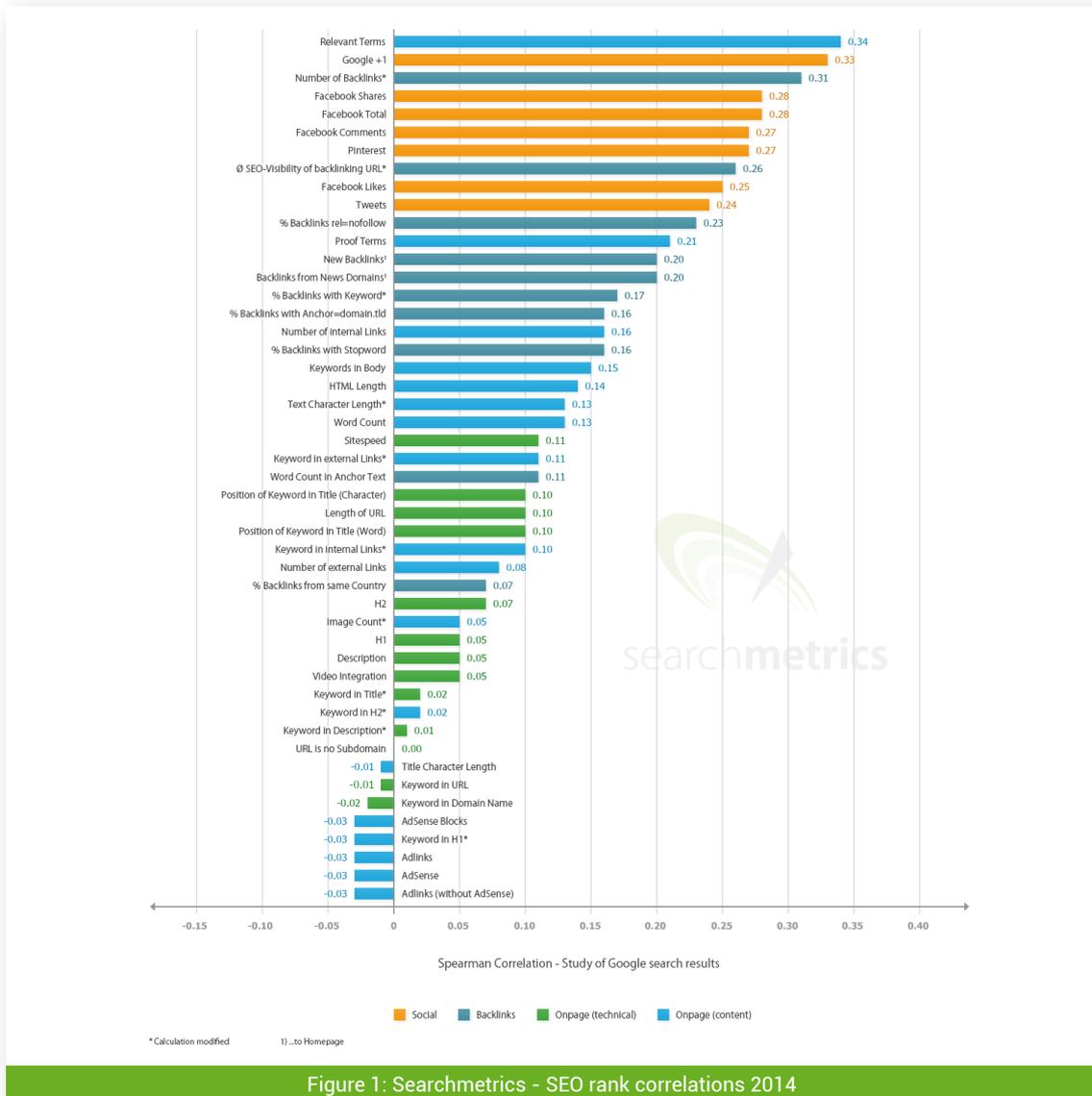


Figure 1: Searchmetrics - SEO rank correlations 2014

Compared to 2012 and 2013, significantly more factors were considered - some are already present in this overview chart, others are covered in the corresponding following chapters. Last year we decided to introduce color coding according to thematic clusters, and we continued to do so in this report.

Explanation: The x-axis describes the size of the correlation coefficient. The longer or shorter the bar in the positive range, the higher or lower the correlation is between that factor and a good Google ranking. Factors with a „zero“ value, according to our analysis, indicate no measurable correlation between good and bad Google results. Values in the negative range are best interpreted by reversing the statement for those in the positive range above.

Binary And Numerical Factors - Specification Versus Existence

The factors examined are divided into binary and numerical features. This means there is an inherent difference in nature of the factors used in this analysis. This difference should not be neglected in the interpretation of the values.

Elements that are described using binary factors - such as a meta description on the page - are either present or not. There are no gradations. There are also elements with graded characteristics that change in differing degrees. For example, a URL may have zero or six thousand backlinks - but each value in between is also possible.

These numerical factors are in some senses „better“ for studies using correlation calculations such as Spearman, with rows and / or ranking principles based on gradations. When it comes to interpretations based on the pure correlation value alone, the statements for numeric factors are often more significant.

To support the validity of the correlations of binary features in our study, therefore, an average value must be specified in the rule. For example regarding the factor „Existence of Description,“ it may be that there is only a close to zero correlation with rankings, but in fact almost 100% of all URLs have a description (see the example correlation for factor A above).

Correlation Values Versus Average Values And Curves

The correlation values plotted as a bar are always calculated on the basis of all available data per feature, so for one or 10,000 keywords, each graph line has 30 values per keyword.

The average values are calculated from these mean curve values, which also leave out the top five percent of each feature to allow for the smoothest possible curve and a manageable scale (Y-axis), otherwise some lower averages in respective curves would not be seen. Also, individual median and mean values are calculated for features disregarding the top five percent of findings.

Adapted Feature Calculations

Some factors relating to crawling and calculation have been improved this year. These features are marked in the bar graphs and the method of adaptation, and the effects on the results, are discussed in each respective chapter. The revisions made have contributed to an improvement in the analysis and evaluation of this year's factors, but are also partly to enable more accurate comparison of results from previous reports.

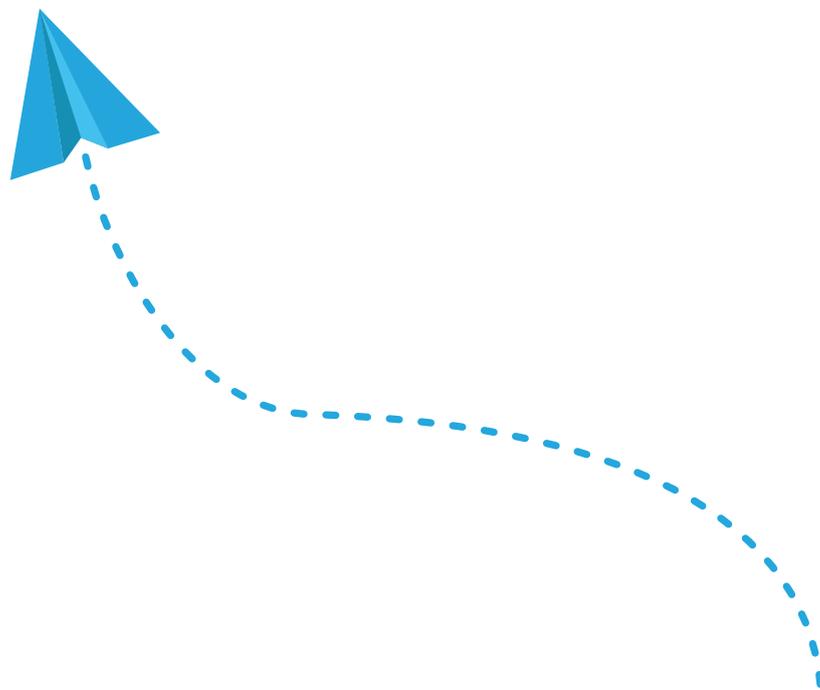
Structure of the White Paper

This study is structured in order to highlight the gradual development of a domain and also to take into account historical developments of relevance factors.

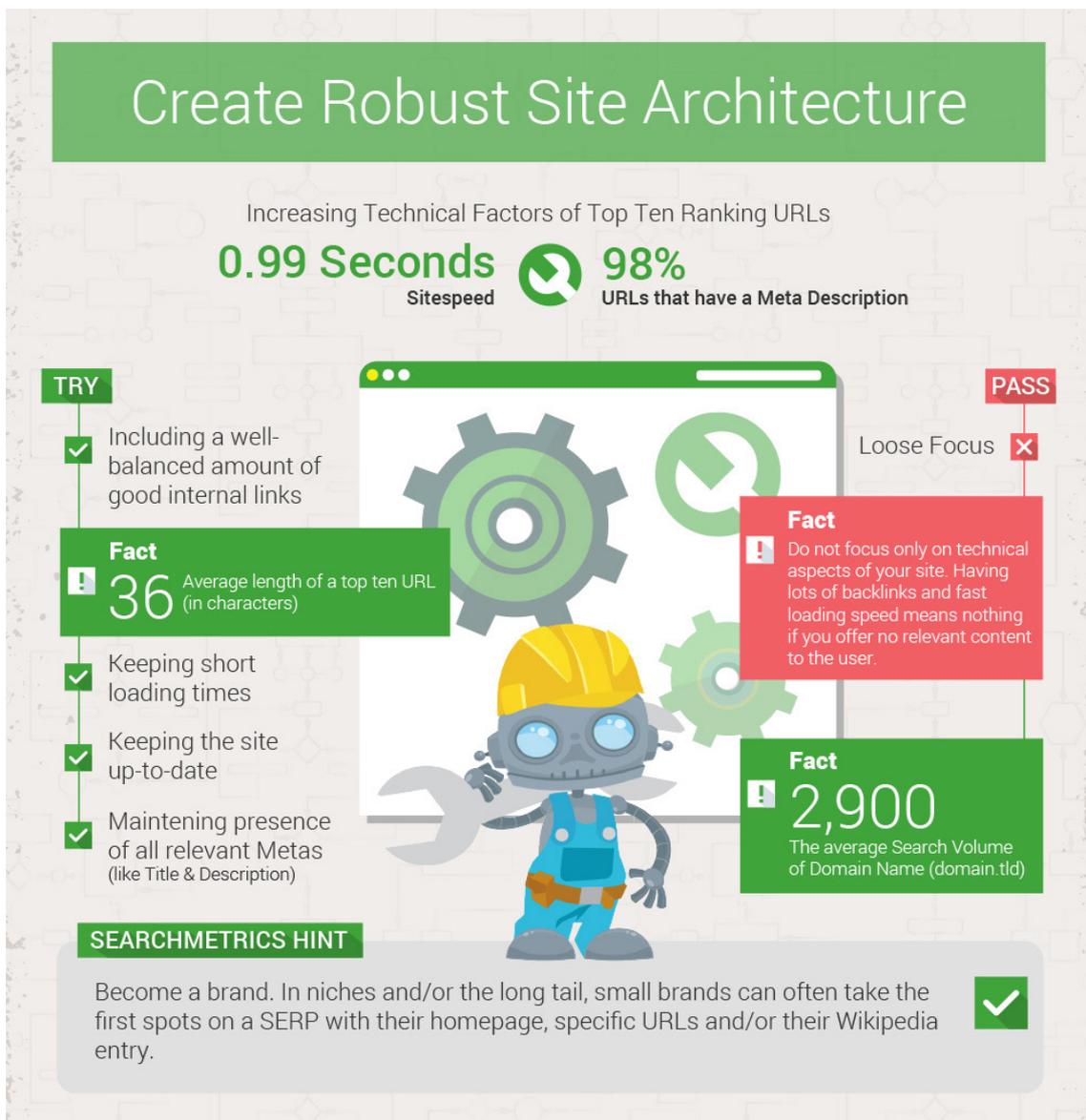
The first section addresses Onpage factors, and is divided into Technology and Content sub-sections. This section is followed by an analysis of complex off-page factors such as backlinks and

social signals. We will conclude this paper with a consideration of the two factors that play a key role in the present studies and therefore require separate versions. Analysis of the relevant chapters on the brand factor and the role of user signals, as well as an outlook on mobile ranking factors is included in the final summary of the results in the overview.

The first chapter reviews all the factors in the overview and examines the trends from the previous year and compares them to 2013. The core of the chapter deals with statements and conclusions regarding the various factors and a conclusion is to be found in the chapter summary.

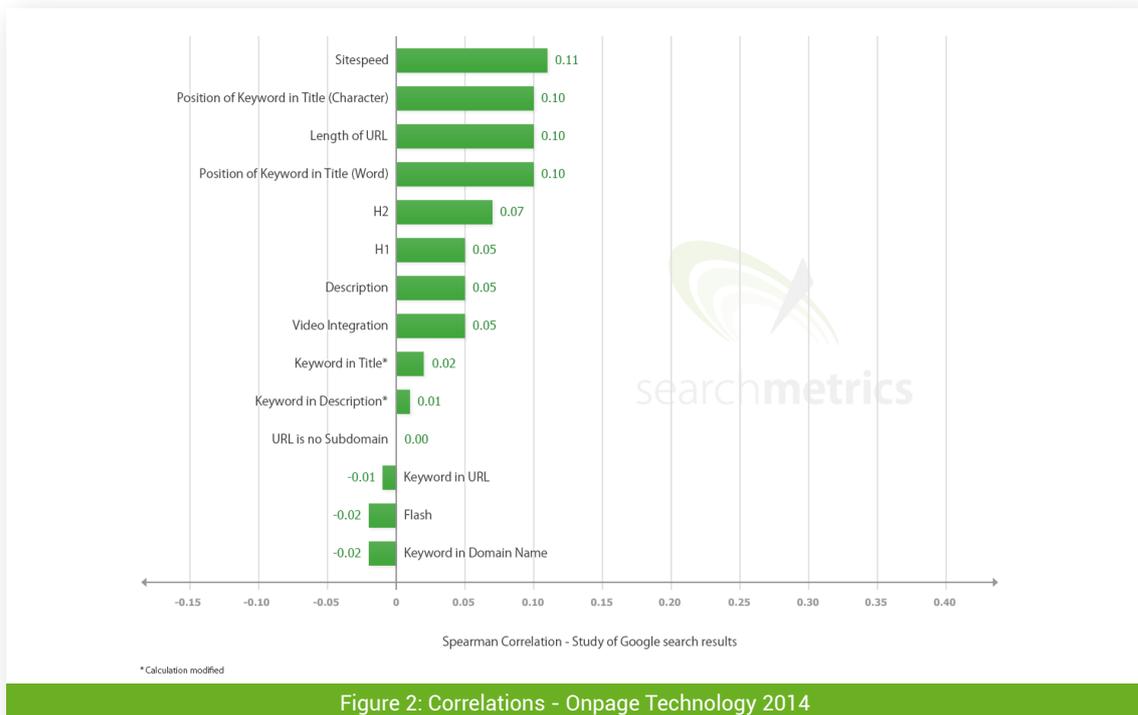


At a glance: Infographic – Technology



1. ONPAGE TECHNOLOGY

Trends noted in previous studies regarding onpage technology have become more pronounced. It appears that it is fundamental for a good ranking to include all onpage factors, especially when technical parameters of the site are concerned. Search engines do not seem to favor pages that meet certain onpage criteria; rather the absence of these criteria have a negative effect on the ranking.



The interpretation of low correlation values, especially with regard to binary features which are common in an onpage technical approach, must be considered carefully. The average value of the correlation should be used for each analysis point.

For the top 30 sites analyzed, the highest ranking pages share similar features. For binary features especially, there is a positive correlation for their presence and a negative correlation for their absence. In other words, if the either/or factors of all the analyzed sites meet specific criteria, no positive rank correlation coefficient can logically be calculated.

It seems that the presence of certain onpage technical factors is a basic requirement for a ranking on the first pages of SERPs.

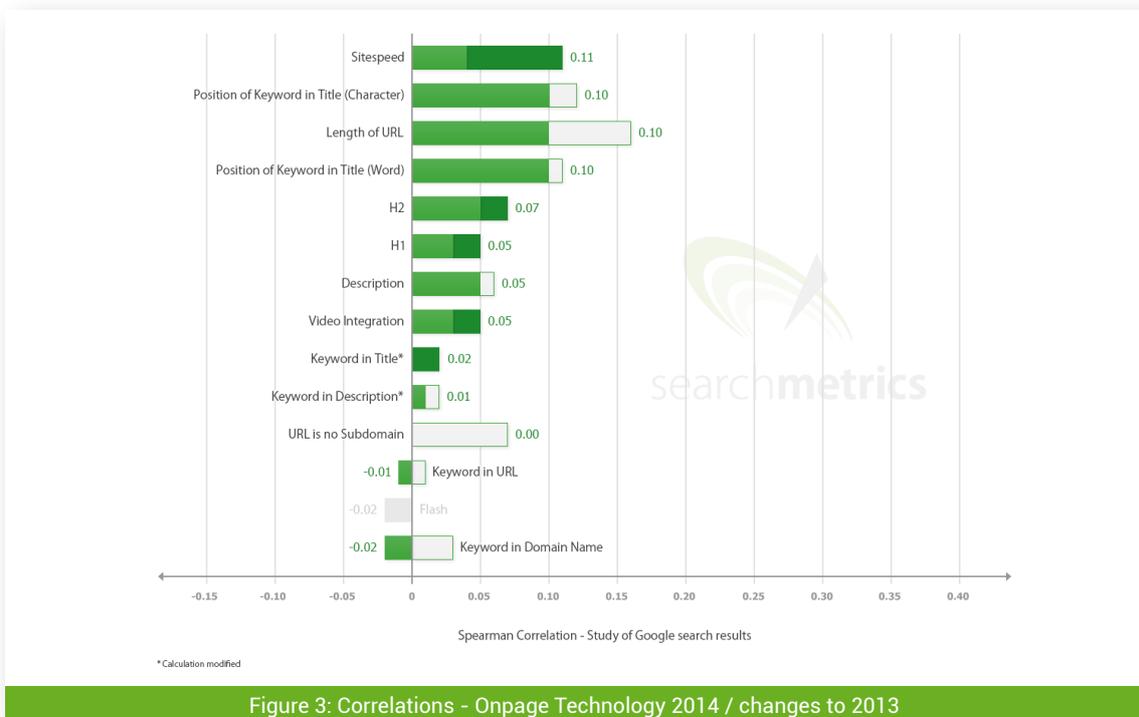
The results from the past year show that this no longer applies to keywords based on the domain name or URL. In 2012 we were able to observe significant positive correlations, but these have now reduced further, after a decline in 2013. The average values continue to diminish here.

The introduction of two new factors last year - „URL length“ and „URL is not a subdomain,“ shows a significant move towards simplicity. Although in both cases the correlation has decreased, the average values are very similar to those from the previous year. Here Google is constant.

These two closely linked features, despite the changed correlations, continue to be interpreted as follows: Shorter URLs and domain names appear more commonly at the top of the rankings. This is especially true for position one, and is explained by the fact that on average (over the data set of all webpage rankings) the primary domain will perform better than its sub-domains or related domain pages.

Onpage Technology: Development Compared To Last Year

The following detailed view of the development of the correlation of each factor compared to 2013 includes new features that only appear in the current study, and for which no basis for comparison exists. Where no comparative data is available from previous years, these figures are grayed out.



In the graph you can see that the correlations for onpage technical factors increased during the year. However, this cannot be taken as a conclusion about the quality of the factors.

While both the correlation for the keyword on the page itself, and the relevance of the keyword position in the title, (the farther forward the better) have increased significantly, the previously significant positive influence of the keyword in the domain name has been devalued again. The feature now has a negative correlation, as well as „keyword in URL.“ This is also reflected in the average values.

Still Few Keyword Domains In Search Results

In 2012, our analysis stated:

„The power of keyword domains has long been known, and is also clearly visible in our analysis.“⁴

And our 2013 analysis stated:

„Those days are over now - the positive impact of keywords in the URL, in particular in the domain name, that had lasted for years, has declined significantly, so much so that keywords in the URL / domain no longer exists as a ranking factor.“⁵

This trend continued in 2014. The following graphic illustrates how the correlations for these two features have evolved within two years.

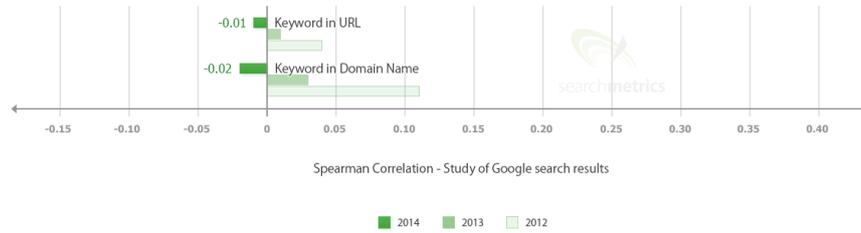


Figure 4: Correlations - keyword in domain / URL

After a considerable decrease in the correlation for „Keyword in Domain“ from 2012 to 2013⁶, this feature has now dropped even further. The correlation for „keyword in URL“ dropped to nearly zero in 2013, and has now changed to a negative.

It is a similar case for the „keyword in domain“ factor. Already declining in 2013, the comparison between 2013 and 2014 shows that this factor has further declined in significance, year on year.

In figure 5, the average values for „keyword in domain“ are presented for a more detailed analysis:

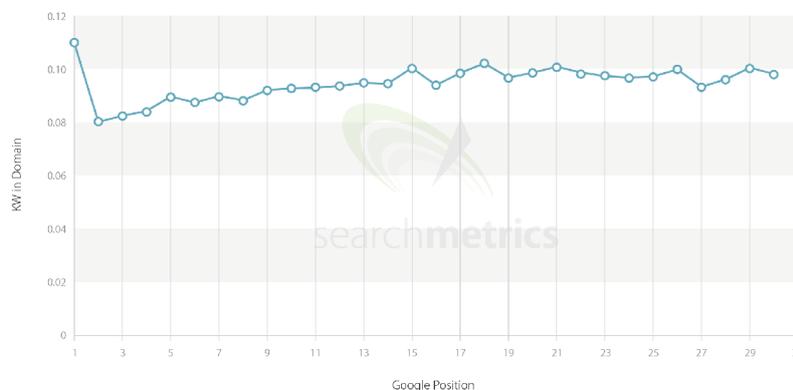


Figure 5: Average ranking - keyword in domain 2014

⁴ cf. Whitepaper - Searchmetrics Ranking Factors 2012 (chapter 5)

⁵ cf. Whitepaper - Searchmetrics Ranking Factors 2013 (chapter 1)

⁶ cf. Summary in the last chapter

The average proportion of domains that contain the keyword in the domain name has declined over the period for the top positions and is in the lower percentage range (y-axis: 1.0 = 100%). While we see a clear decrease compared to 2013, it's not severe. The only exception is a slightly higher proportion in position 1, without which the correlation would be even more negative.

The curve for „keyword in URL“ looks very similar, except that the mean values are significantly higher than those shown in the last graph. The reason for this is likely because domain names are limited and are subject to fixed registration. URLs have much more variety with much less effort, and appear more often with the keyword included. Approximately 39% of all URLs in the top 30 have the keyword in the URL.

In the following section we address some factors that demonstrate that it is not a simple task for Google simply restructure the field.

Keyword Frequency In Title And Description

By now it is well known that over-use of keywords has a negative effect on rankings, but there are some onpage areas where the existence of keyword factors generates a positive correlation with good ranking, with average values for 2014 higher than for 2013. For onpage optimization, the presence of keywords plays a significant role.

All „keyword in X“ features have been slightly adjusted this year. The rules on recognition of the keywords in an HTML document on a URL have become more accurate, but also more stringent. Now multi-word keywords are being recognized more efficiently, even if the individual elements are separated by stop words.

The following graph focuses on this area, and the two related content factors are included in the chapter below for clarity. All chosen factors have been revised for this study about the recognition of keywords.

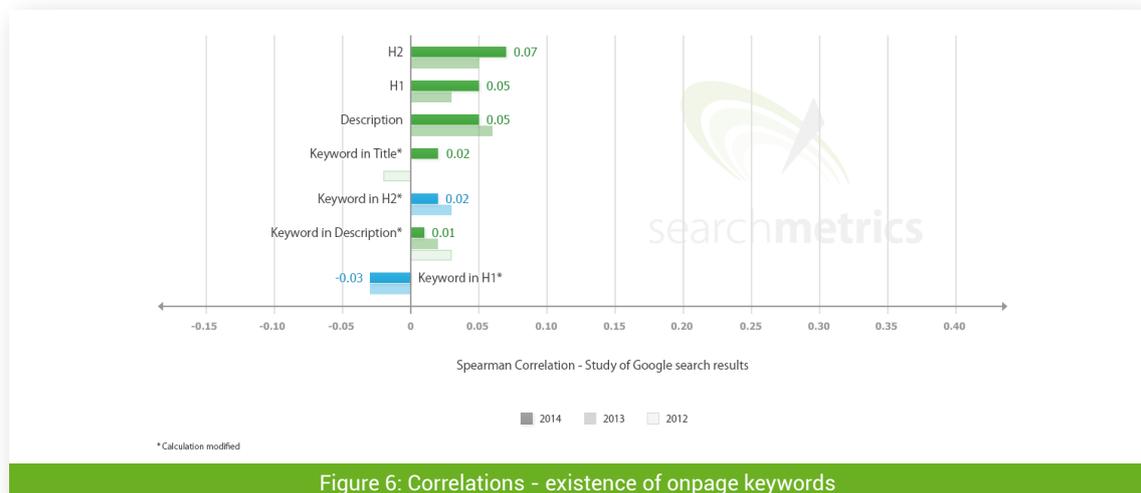


Figure 6: Correlations - existence of onpage keywords

Onpage keywords will not disappear. Google is a machine that structures its data according to the relevance of search queries, and it is moving closer to its goal of Semantic Search, especially with the Hummingbird update. The process of analyzing content (read: words/text) for semantic reasoning (read: meaning) is the pinnacle of decoding. But, until then, the presence of onpage keywords within content will continue.

However, the days of „Keyword Density“ are over. Apart from the brand-factor (which does play a role), it is, and always has been, a positive thing to include keywords in the description, the title and in the H1 heading. If you compare this to a book, for example, if the title and description isn't attractive, you won't read the book. Titles and Descriptions that the user reads in the SERPs snippets also encourage clicking, or not.



Figure 7: Average ranking - existence of keywords in Title / Description

The graph in figure 7 shows that about seven out of 10 pages include the keyword in the title, but amazingly, only slightly more than half use the keyword in the description. In the top positions the brand factor can be seen, and this is significantly more pronounced in the Meta Description. Brand domains have the keyword in the title less often and even more rarely in the description. The difference between the URLs that have risen to the second half of page one, and the URLs at position 20-30 is greater than the factor „Keyword in Title.“

In particular, Wikipedia is often well ranked even without keywords because Google understands the relevance of Wikipedia articles, and they are also significantly longer than the norm. Partial queries can then be rewritten internally for these purposes by search engines, as with the keyword „jt.“

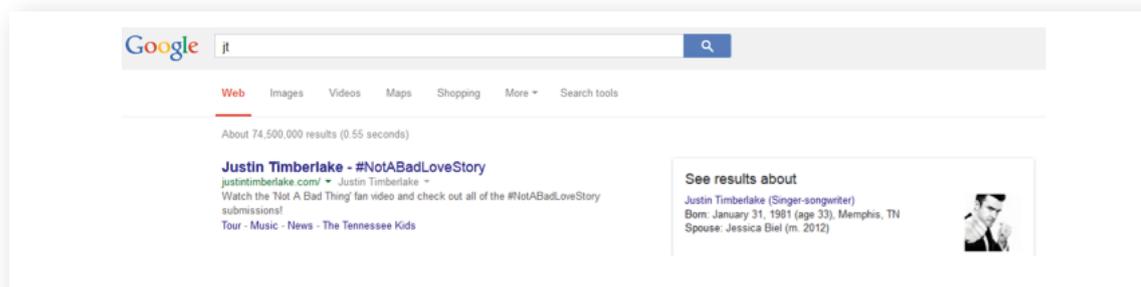
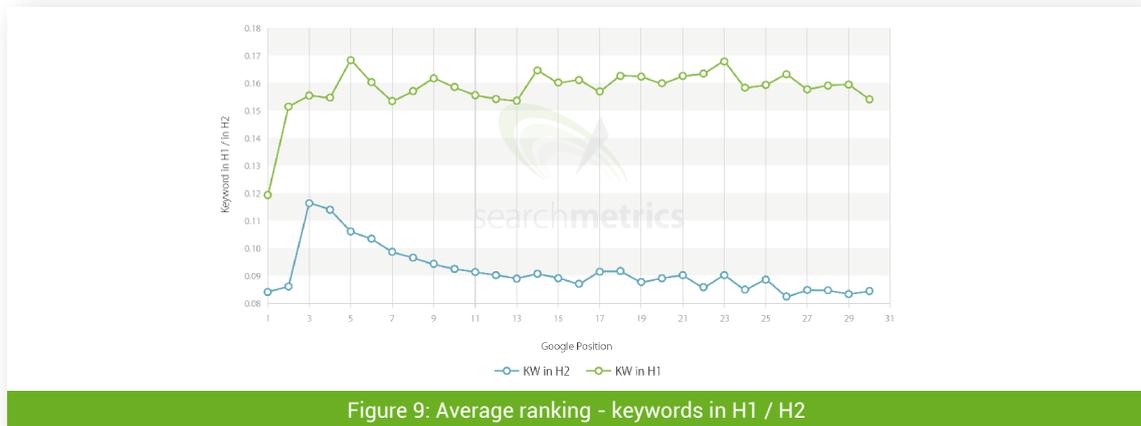


Figure 8: Google SERP for keyword "jt" - Screenshot

This search query recognizes Justin Timberlake's domain as a brand (without having the keyword in the source code), and on the right is the explanation, linked to the Wikipedia entry at position 2. Notice that both snippets appear without the entered keyword in the title or description.

Content Digression: Keywords Are Rare In Hx-Headings

Because these features are closely linked to the content of a page, let's take a quick look at the corresponding Content-Feature „keyword in H1/H2“ whose correlation values have already been included in the detailed bar chart above.



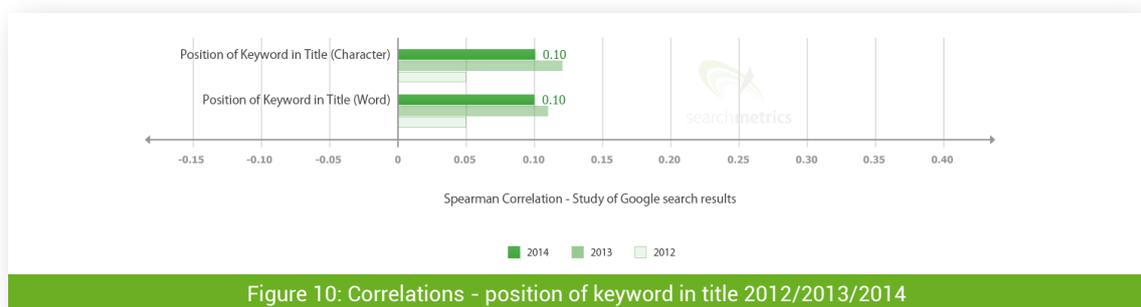
It can be seen that the use of keywords in headings is rare in the top 30 positions. Once again, Brands rarely place a keyword in H1 or H2, yet still rank well for them. Because Google has a high interest in the most relevant results – often Brands – it prioritizes this placement (see the Justin Timberlake example).

It also appears that keywords on the main page are more important than those in H2 headings.

Position Of Keyword Quite Early In The Title

Last year, keywords in the page title and description showed positive, even growing, rank correlation coefficients. The correlation in respect to position of keyword in title had risen from 2012 to 2013, but this year we see a slight decrease. Does this mean keyword in title has become less important? No. It just means that other factors gained more importance in relation.

The following graph shows the Spearman correlations for the values „position of keywords“ for both words and characters. These findings can be summed up as – ‘the sooner, the better.’



The graph shows the development of correlations over a three year period, from 2012 (bottom bar) through 2013 (middle bar) to 2014 (top bar). The positive value for both correlations, which are also closely related to each other, has declined slightly after doubling in the past year. The differen-

ce between better and worse ranked pages is, consequently, decreased slightly. The top positions of the search results continue to return many pages where the keyword in the title is placed further forward in the title tag.

But here, looking at the averages, it appears that keywords in 2014 had not been placed further forward in the title than last year, especially the on first page of search results:

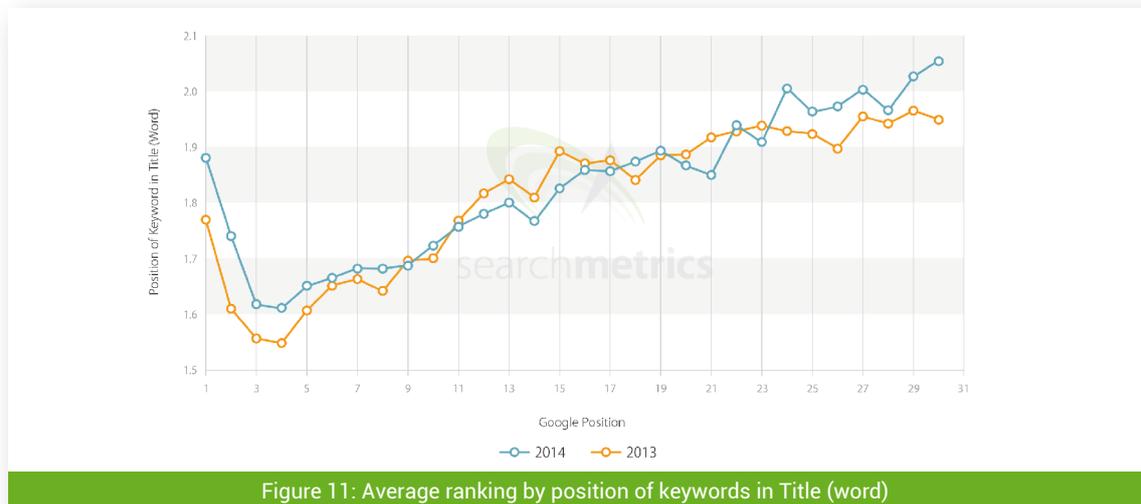


Figure 11: Average ranking by position of keywords in Title (word)

Page Load Times As An Important Performance Factor

One feature that is highly technical, and that was available for the first time in last year's analysis, is „Load Speed.“ This factor was examined with the expectation of a significant positive correlation in the pool of ranking factors in 2013. However, the correlation for this value, as the overview diagram shows, is relatively low.

This may be related, among other things, to the fact that a fast load time in the top Google positions appears to be trivial. However, last year showed that a shorter loading time in the top positions should have resulted in a positive correlation

With faster loading time sites in the top positions last year, the curve should have shown a positive correlation.

As mentioned in our Ranking Factors study from 2013, however, the Wikipedia results at positions 17 and 27 in the correlation calculation had the effect of flattening the average curve. This year, both the correlation and also the mean curve were calculated without including Wikipedia, and the correlation is positive.

Indeed, the Site Speed for Google is measurable directly, while crawling pages, or indirectly via user signals such as CTR⁷ and Bounce Rate, and is most likely also a ranking criterion. Slow loading pages simply have poorer user signals.

In the detailed view (below) it's clear that the top 30 pages load quickly, with sites ranked first on average up to 15 hundredths of a second faster.

⁷ Click-Through-Rate, cf. chapter 6

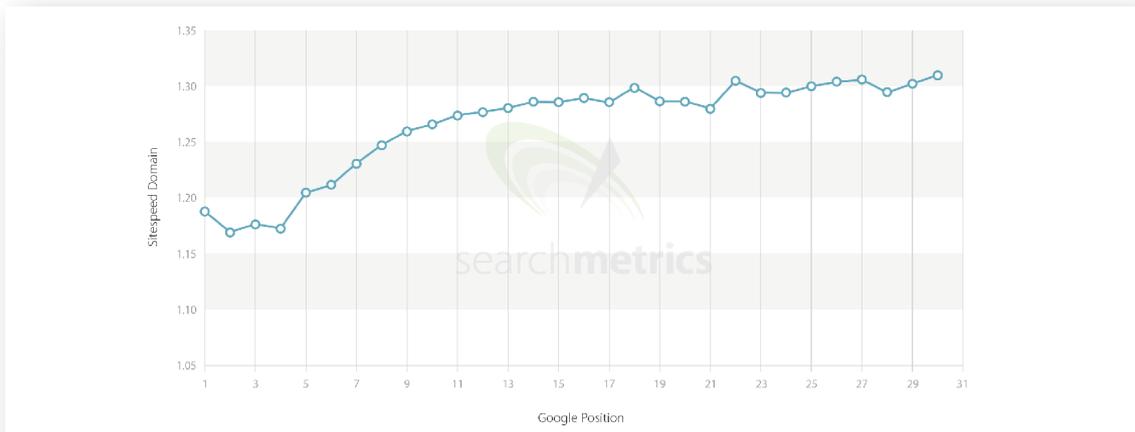


Figure 12: Average ranking - average page load time, domain URLs

The average load time of a domain - average load times of URLs on the same domain - in the top 30 is 1.27 seconds.

It seems surprising at first to note that URLs seem to load a little slower on average than the previous year. One reason could be increased file sizes - as will become clear later - related to other content features such as text length. Furthermore, as shown in the above graph, the load speed values for URLs in the same domain are averaged.

Looking purely at the URL level, without extrapolating the average load speed per-domain, the following picture emerges:

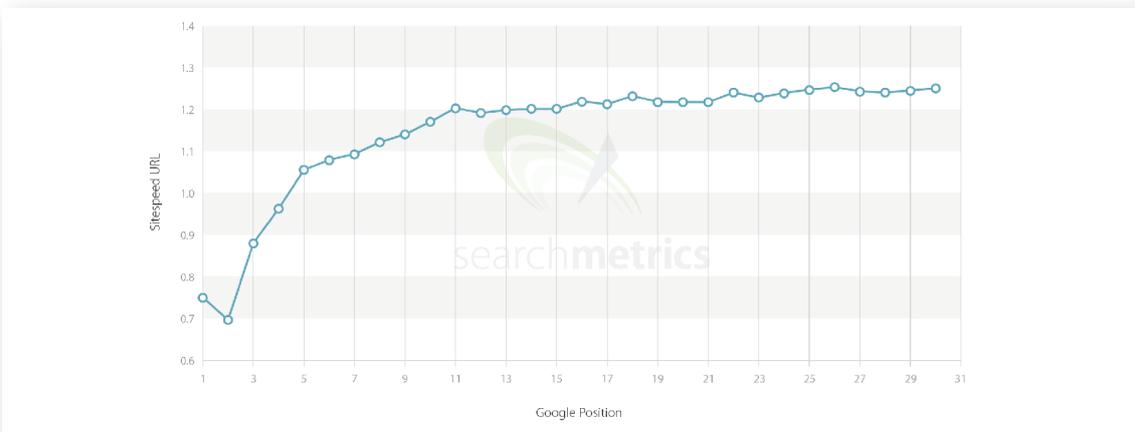


Figure 13: Average ranking - page load time per URL

URLs at position 1 load on average about five tenths of a second faster based on this calculation - faster than both the rest of the top ten positions, and faster than URLs at number 1 in the average calculation.

One possible explanation for this would be the presence of Content Delivery Networks (CDN). These connect via the Internet at station server networks (SSNs), and are used to optimize user-query performance. Data (the most frequently used content) is pre-cached and can be delivered faster and more efficiently for user requests.

Use Of Flash No Longer Widespread

Given the importance of website loading times as a ranking factor, we also queried the widespread use of Flash. This analysis is new for 2014 and was introduced as a response to the introduction of mobile coding via HTML5 etc., and is most interesting for desktop users.

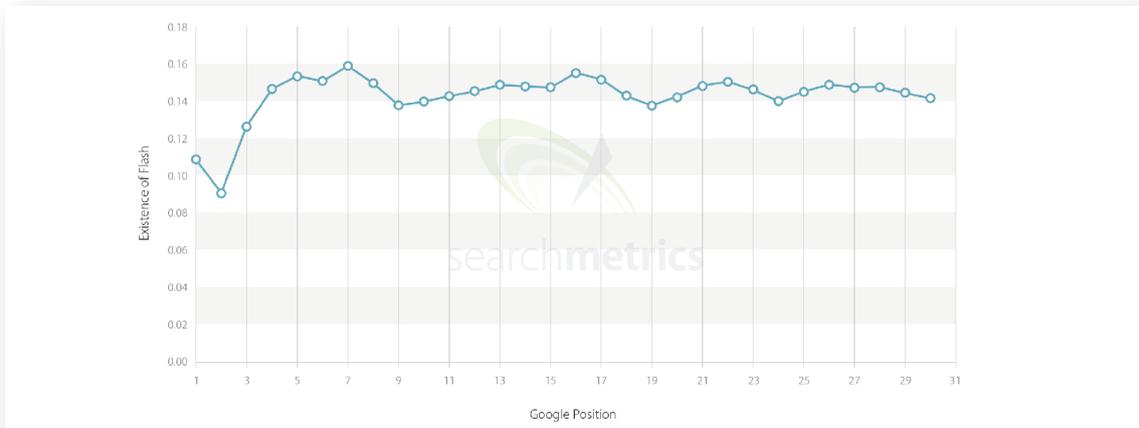


Figure 14: Average ranking by Flash

The correlation is -0.02 with almost zero and, at first glance, it seems to make no major difference to the results. However, upon closer inspection, the Brand factor can be recognized here, as well as a curve that flattens slightly towards the end. On average, only 14.27% of the analyzed sites still use Flash today.

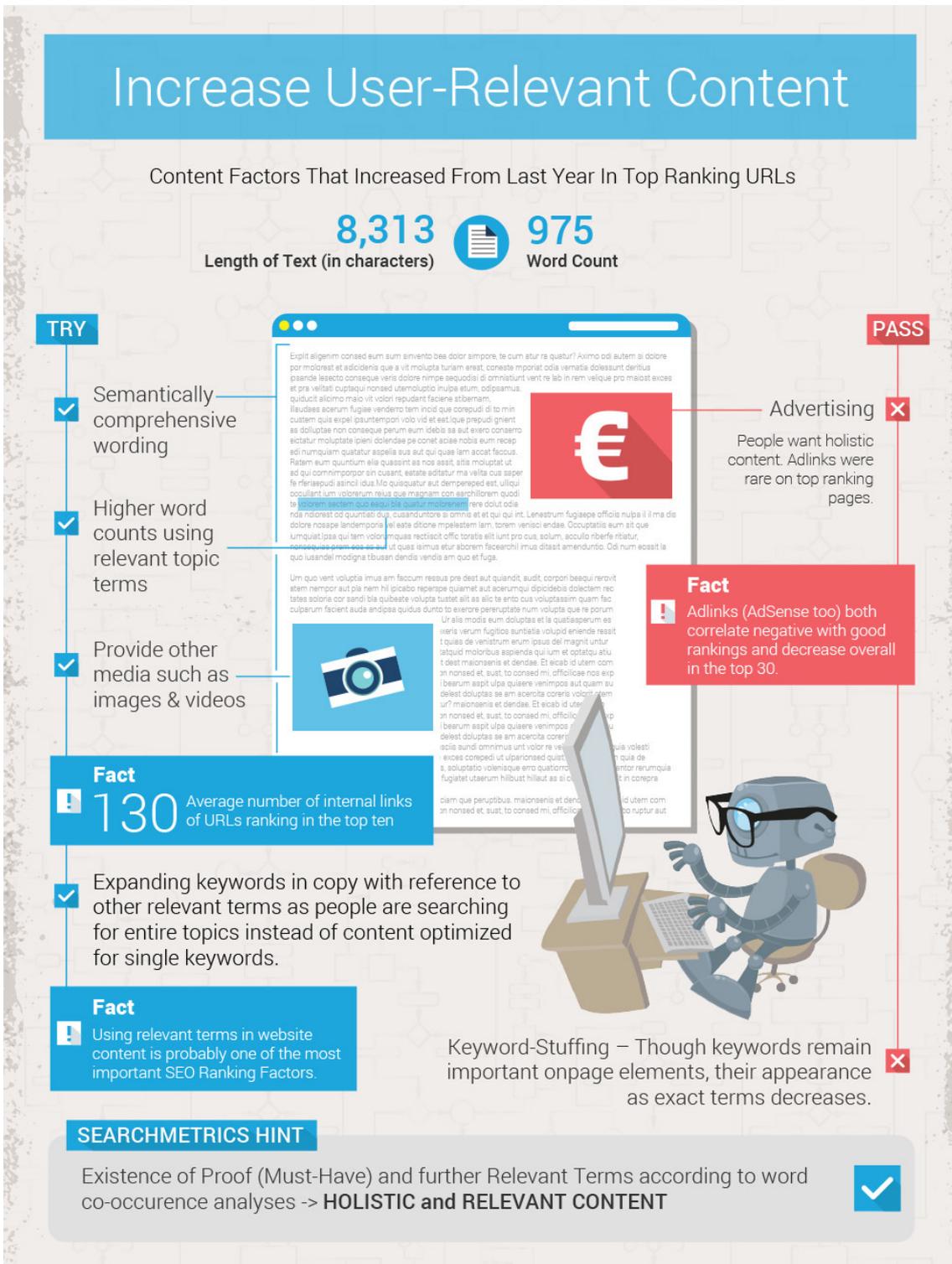
Summary

The differences between positions 1 to 30, and hence the correlation values, are slight in terms of the technical criteria. This is because the implementation of technical on-page factors is relatively simple and a prerequisite for ranking in Google's top positions - a strong technical site structure and performance is one of the absolute SEO Basics.

The factor „keyword in domain,“ and keyword domains as a ranking factor, were devalued in comparison to 2013 for the second year in a row. The presence of keywords in the Title, the Description and in H1 and H2 tags is still best-practice, but should not be over-optimized. A quick page load time is one of the basic requirements for good rankings. The top pages load, on average, significantly faster than the rest.

The presence of H1 headings has risen on average compared to the previous year, but still only about 75% of all top 30 URLs have them, and only about 69% of sites have at least one H2 heading.

At a glance: Infographic – Content



2. ONPAGE CONTENT

Content, and especially its quality, is becoming increasingly important. This was not the case for a long time. In the past, keywords dominated the SEO market, and they will continue to be the basis for search. However,, for some time now there has been a shift in emphasis away from pure keyword strategies towards more complex topics. Some other factors have diminished in favor of content.

Content, as well as the technical organization of a site, is directly influenced by the site editor, but has often been subordinate. From an SEO perspective, other areas were often optimized - the content, at least from a qualitative point of view for the user, was usually not. This was due in part to the evolution of the search algorithm, which was nowhere near as advanced as it is today. With Hummingbird, Google has set new standards in the areas of semantics and context, which can be monitored in the SERPs⁸.

Content features are often complex and operate on a sliding scale determined by Google, with 'Relevance' as the ultimate goal. Google wants to evaluate and structure content according to its quality and relevance to a search enquiry.

For this analysis of the content features, the number of factors analyzed over the previous year has been significantly expanded.

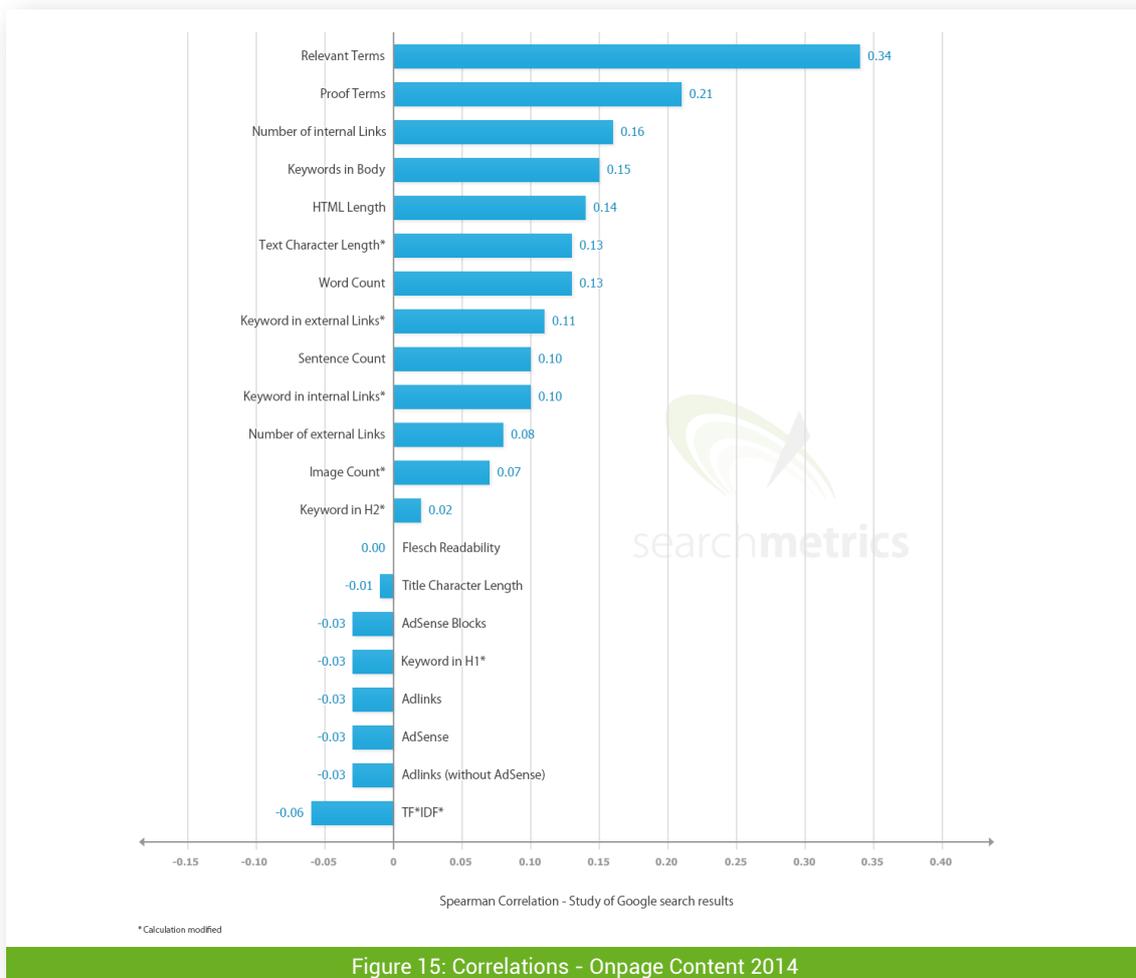


Figure 15: Correlations - Onpage Content 2014

In principle, content factors can be divided into three groups:

- | Composition and quality
- | Linking
- | Advertising material

With some exceptions (discussed later in the study), content factors are positively correlated with good rankings. In the overview of previous years' trends, the relevance of page content is evidently important.

Onpage Content: development compared to last year

The graph below shows the development of content features that were part of last year's study.

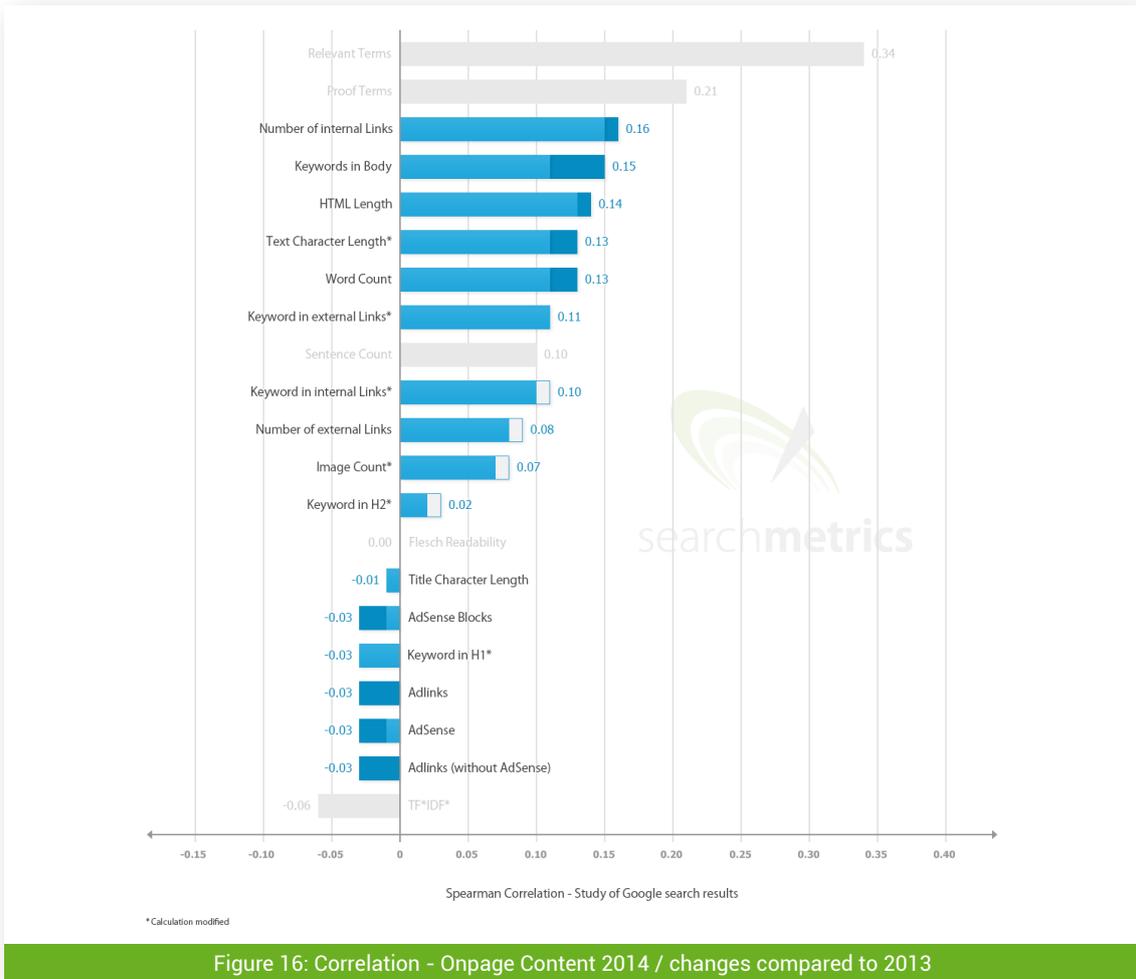


Figure 16: Correlation - Onpage Content 2014 / changes compared to 2013

It can be seen that the majority of factors have increased in importance, sometimes significantly. On the other hand, advertising seems to have decreased.

Although „Keyword in H1“ and „Title length,“ are negative correlation factors, websites that rank better have shorter page titles on average.

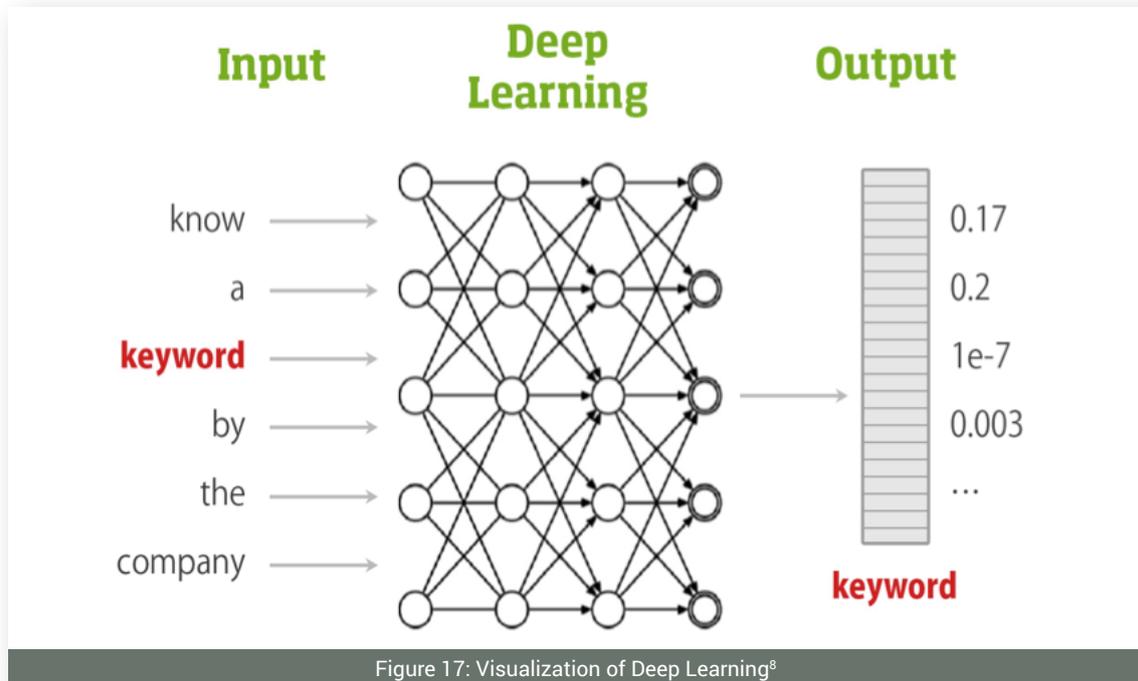
2.1 CONTENT QUALITY - NEW FEATURES

Last year’s analysis showed more importance being given to content features. However, TF * IDF was not enough for us then. The Searchmetrics view of the effects of Hummingbird showed the diversity of search results decreasing, particularly when comparing sites in SERPs for related keywords with overlapping search intent.

This is based on semantics and context. For example, it is very likely that the word „car“ is relevant in a text in which the word „bumper“ occurs, while the same is not true for the term „refrigerator.“ This is based on Co-Occurrence analysis.

„You shall know a keyword by the company it keeps“ (Firth, 1957)

Google is good at analyzing the semantic meaning behind keyword strings. The development of search continues to move away from the meaning of a single keyword towards analysis of so-called 'content clusters' – individual subjects or topic areas – that can be based around keywords and developed naturally into content copy. The difficulty is then to make this machine-measurable.



There are, of course, different calculation methods to analyze semantic relationships between terms.

All these calculations result in the following statement: The more topics a text reflects, the more holistic it is – and, at the same time, the more relevant it is for users with different search intentions. Consequently, the copy also ranks better for related, additional keywords as well as the primary keyword.

A notable feature is that „stop words“ – definite articles (the), prepositions (to, after, during, etc.), etc. - are filtered out of the analysis because they are neither important nor relevant.

If website editors want their content to rank better for specific keywords, the content should be created with the fulfillment of user search intent in mind. While this should go without saying, for a long time it was technically difficult to optimize for.

Content Clustering With Word Co-Occurrence Analyses

One way to make semantic relationships between words and text relevant and measurable to a particular keyword is to examine the copy in terms of Co-Occurrence approaches and 'Proof Terms'.

Proof terms are essential. The semantic relationship to the primary keyword is very high when these terms are present in all documents relevant to the topic and are regularly used. Example: Main Keyword: Panda Update / Proof Terms (examples): Google / Panda.

⁹ Graphic (©Searchmetrics) based on <http://www.cs.toronto.edu/~rsalakhu/isbi.html>

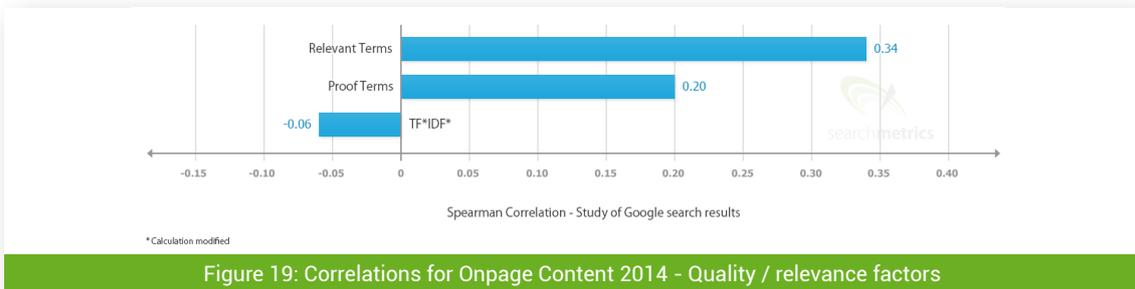
Individual keywords (1 to 25 of 250)

Keyword	Relevance	Documents in %	max. weighting
✓ google -	41.57	100.0%	0.60
✓ panda -	40.55	100.0%	0.59
webmaster -	22.03	88.9%	0.29
rankings -	19.22	88.9%	0.25

Figure 18: Content Optimization Searchmetrics Suite (screenshot cropped)

Relevant terms, however, are semantically removed relatives of the main keywords, and are usually part of a subordinate topic cluster. They are not mandatory but are often included in the main copy. Example: Main Keyword: Panda Update / Relevant Terms (examples): Webmaster / rankings.

The correlations for these features are as follows:



For comparative purposes, the correlation for TF * IDF is included above, and, as in the previous analysis, it's negative.

A brief explanation of 'TF * IDF': The relevance of a term in a document, and its relevance for a keyword search should, according to this approach, be applied not only to the individual web page in question, but also compared to the total number of other pages with this keyword in the index.

The correlations for Proof and Relevant terms are very different from those for TF * IDF with respect to the mean values, as follows:



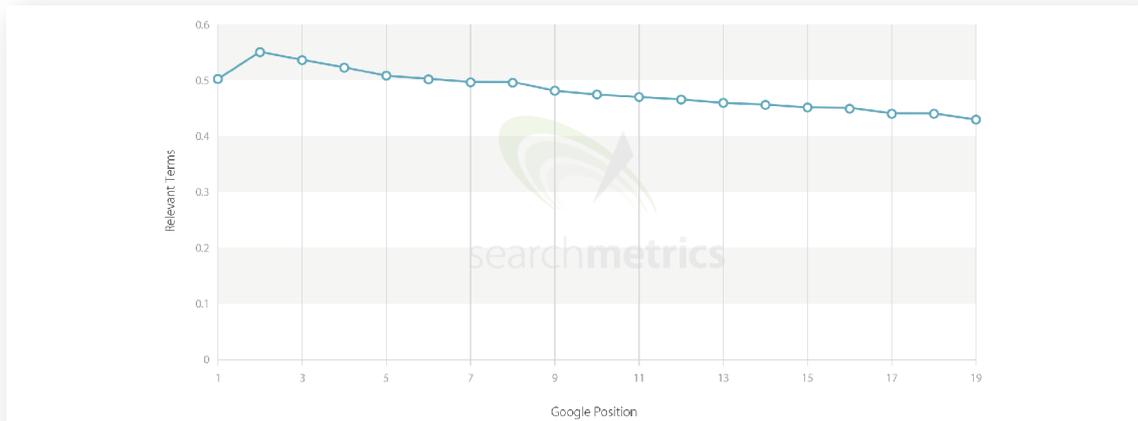


Figure 21: Average after ranking - Relevant Terms

For both of these features, the Brand Factor comes into play, producing a declining curve that shows itself in the correlation values. The scale for the „Proof Terms“ factor is more detailed in order to visualize the curve better. The correlation for the „Relevant Terms“ feature is significantly higher in this year’s analysis, and is the correlation with the largest total value of all factors studied.

Important features of onpage content for search engines include term frequency, the frequency of other relevant and related terms, the space between keywords and the semantic indexing, but never the number of keywords themselves. On the contrary, if keywords are used too frequently, this is swiftly recognized as spam.

Readability Of Text - Flesch & Co.

In another new development this year, the legibility of the text was analyzed. The results here offer us a study in negative correlation.

To interpret these values according to their style, on average, higher-ranking texts are somewhat easier to read. In the following graph, the scale on the y-axis is the Flesch-value for the English language, based on the following formula:

$$206.835 - 1.015 \left(\frac{\text{total words}}{\text{total sentences}} \right) - 84.6 \left(\frac{\text{total syllables}}{\text{total words}} \right)$$

Figure 22: Flesch Formula for Legibility - English

The formula for other languages differs from this, paying more attention to longer / shorter syllables, grammar etc.

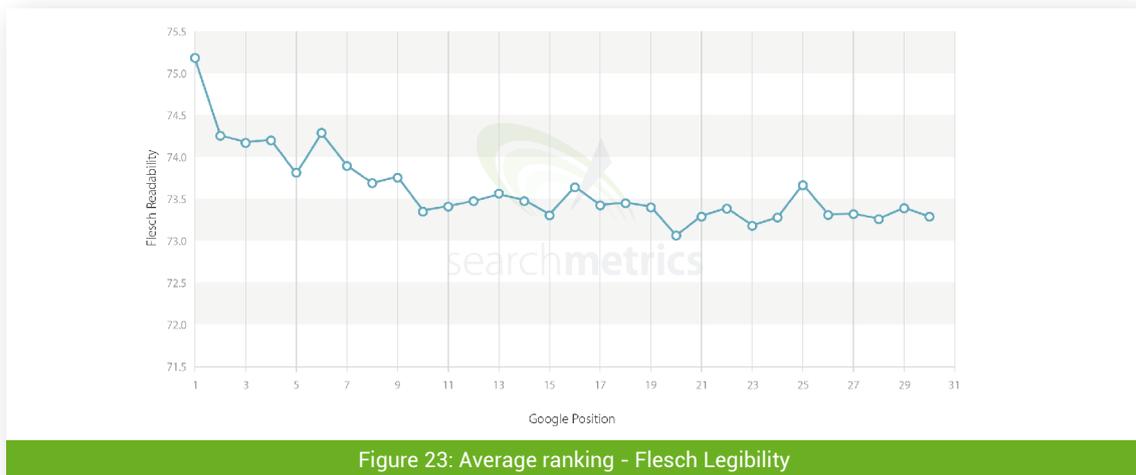


Figure 23: Average ranking - Flesch Legibility

What is interesting here is that the curve for U.S. Google.com search results is fairly characteristic. The correlation is 0.00, but content on the higher ranked URLs tends to be easier to read.

The Flesch-value for the legibility of content from U.S. search results is about 73/74 which corresponds to a relatively simple level of comprehension, as the following table shows.

Score	Notes
90.0	easily understood by an average 11-year-old student
60.0-70.0	easily understood by an average 13- to 15-year-old students
0.0-30.0	best understood by university graduates

Figure 24: Flesch-scale¹⁰

Even Google's Head of Webspam, Matt Cutts, has stated that content that is easier to comprehend by a much wider group of readers will rank higher¹¹. In fact, the determination of the readability of content on search results is probably influenced by the metrics of user signals such as Time on Site and Bounce rate.

2.2 Content length increases significantly (again)

The importance of all the following features has increased for every position in the search results, not only for the first one or two rankings. This means that websites need to produce more content in order to remain competitive in search.

More content is not the same as better content. Text may be subject to several quality criteria, but usability and relevance are always the main goals. To rank higher, sites must produce not just more content, but more relevant, holistic content.

Number Of Words In The Text

The „Number of words in the text“ feature still had a negative correlation to good rankings in 2012. In 2013 this was significantly different, and it has evolved even further this year. This graph shows

¹⁰Source: http://en.wikipedia.org/wiki/Flesch-Kincaid_readability_tests

¹¹ <http://www.youtube.com/watch?v=F4qKcKM7TsE>

a selection of content quality factors, their correlations and their development over the last three years:

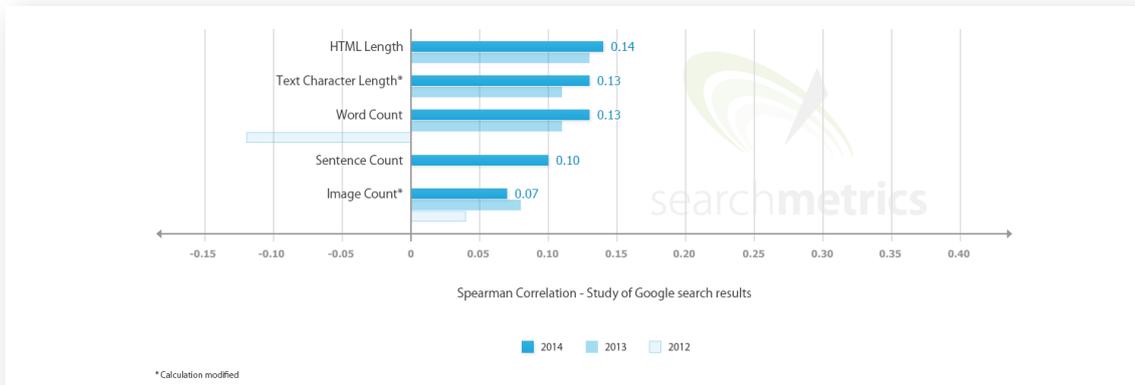


Figure 25: Correlations - content quality 2012/2013/2014

The following detailed chart shows that sites with more words in the copy occupy higher-ranking positions, and sites further down the SERPs have fewer words in their copy. The Y-axis shows the average number of words, the X-axis shows the Google position.

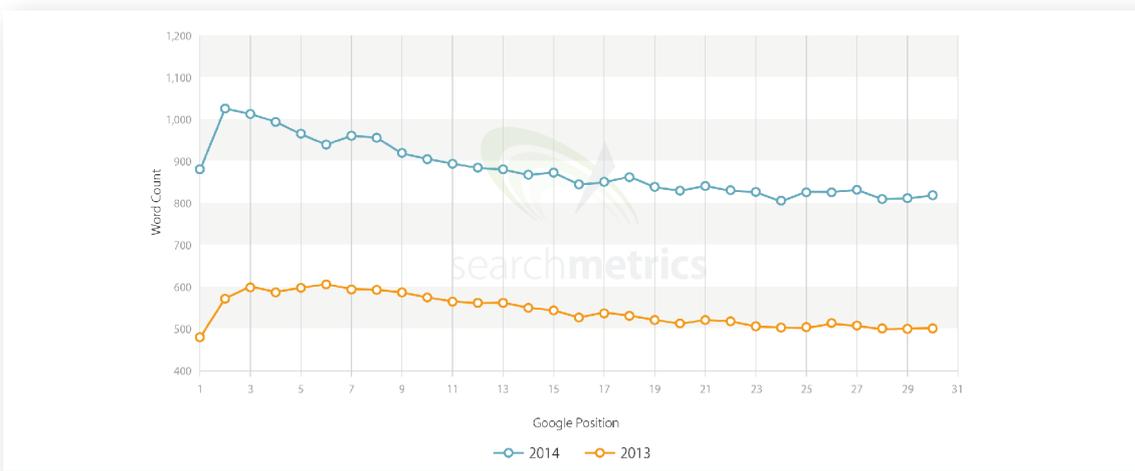


Figure 26: Average ranking - number of words in text 2013/2014

The number of words in the text has risen significantly again compared to the previous year. It can be seen that content is getting longer on average, and continues to do so with the brand factor. URLs in positions two to nine have most words in their copy - more than 900 per section.

HTML Length

Because the length of the HTML code of a web document is in some respects a technical factor, the „length“ factor in this context is highly dependent on the nature of the content.

The vast majority of web documents in the top 30 SERPs have a certain (minimum) length of HTML code. The average of over 300,000 HTML documents evaluated, resulted in the following ratio of HTML length to ranking (Y-axis: average number of characters in the HTML code):

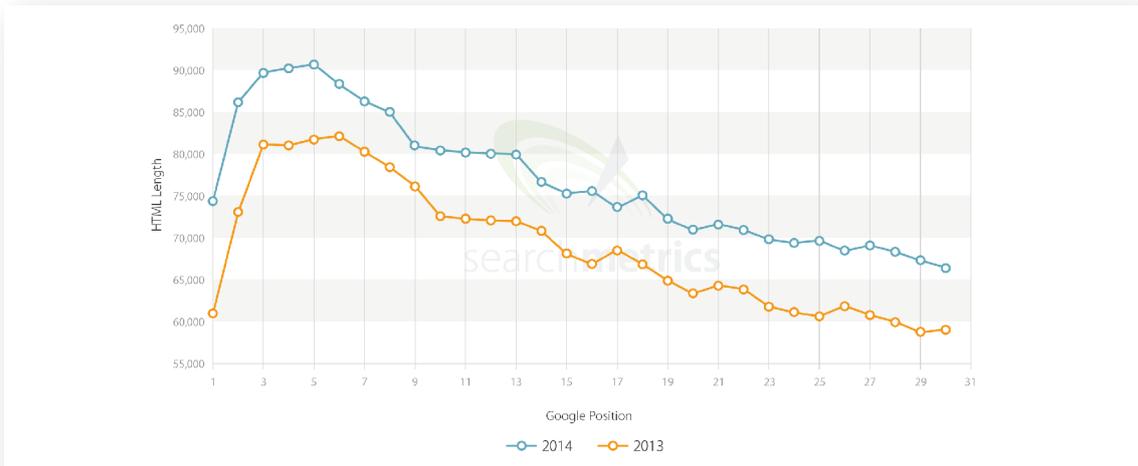


Figure 27: Average ranking by length of HTML code in characters 2013/2014

Again, the Brand Factor can be noted in the top ranked positions, even more than the sheer number of words in the content. From position 5, an almost ideal curve is seen, showing that better ranked pages have longer HTML code than pages further down the rankings. The average length for HTML code in 2014 is 77,169 characters.

It could be observed from this graph that the more HTML characters in the code, the higher the ranking is for that site. This is not true – at least not always. It should be noted at this point that, after about 90,000 characters, the optimum length is exceeded. Larger values of HTML characters do not negatively affect the rankings, but the effect of adding more code above this level ceases to have an impact on this factor. This level seems to vary – last year there was a very similar effect produced for fewer characters in the HTML code.

Text Length

The shape of the curve for the 'Text Length' coefficient is almost identical to that for the HTML length factor. This makes logical sense, as the two factors are co-dependent.

However, the decoding of characters has changed compared to the previous year, which affects the absolute number of text characters. This means that 2014 data cannot be compared to that from 2013. Subsequently, the data for 2013 were recalculated using the current basis, and we have provided a comparison below.

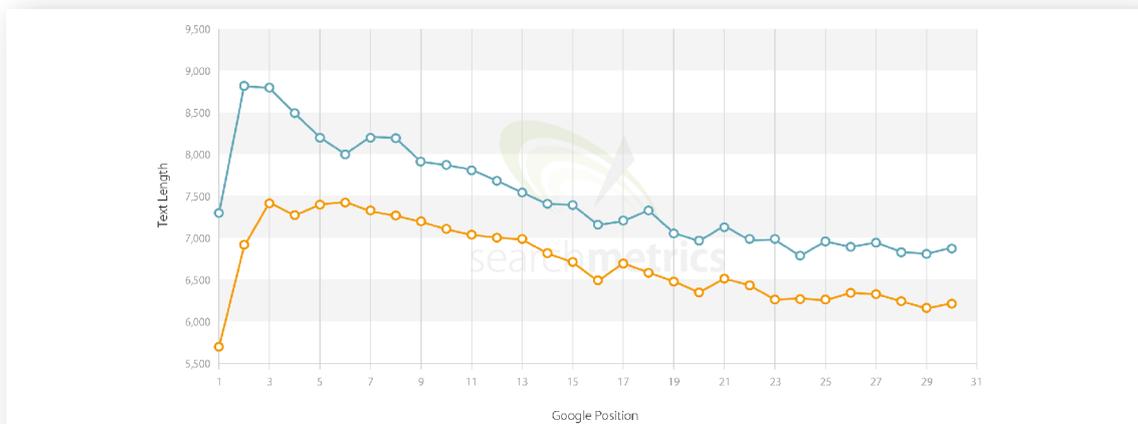


Figure 28: Average ranking according to length of the text in characters 2013/2014

Text length seems capped with respect to its positive correlation with good rankings. Up to this limit, however, excluding the Brand Factor, we can say that pages with more text rank better.

Number Of Sentences

For the sake of completeness, the number of sentences was also examined this year. The feature correlates positively, which was expected from the previous analyses.

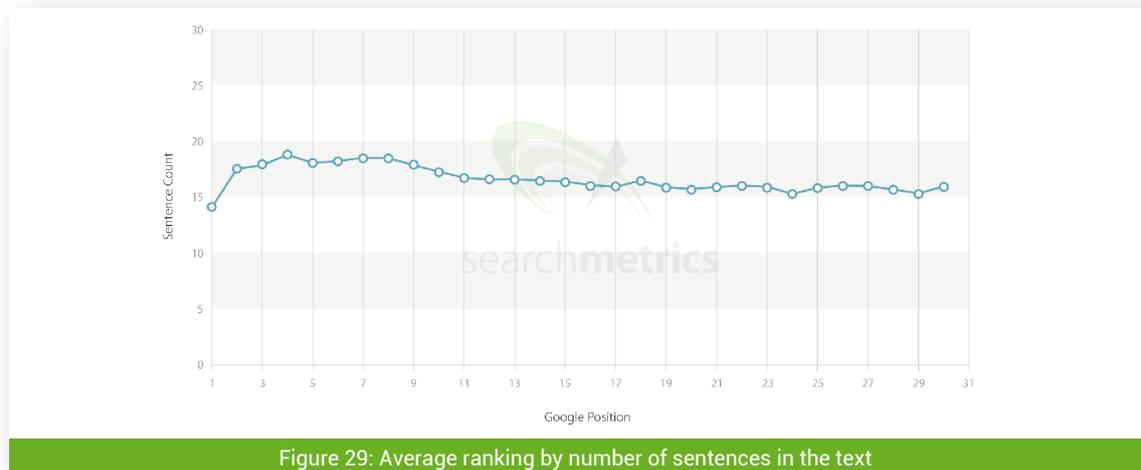


Figure 29: Average ranking by number of sentences in the text

This also shows that the content of Brand sites consists of fewer sentences. On average, the content of pages in the top 30 positions contains about 17 sentences.

It should be noted once again that this is average values and the content analysis can be affected by several navigation or menu elements.

2.3 Enrichment of content by media pays off

With respect to this feature, we have adjusted our analysis and installed a size-filter. The recognition of images is now limited by the filter, and comparisons to previous years, in which both the correlation and the average number of images increased, is no longer feasible.

Photos and videos not only make text more attractive for users, but for Google, this trend is likely to develop positively and be capped at a certain level.

The following graph shows the average number of image files (all files with an image tag on the page) shown by page rank. Please note that an image was considered to be an image if there was a minimum IMG tag with 32px height x 32px width.

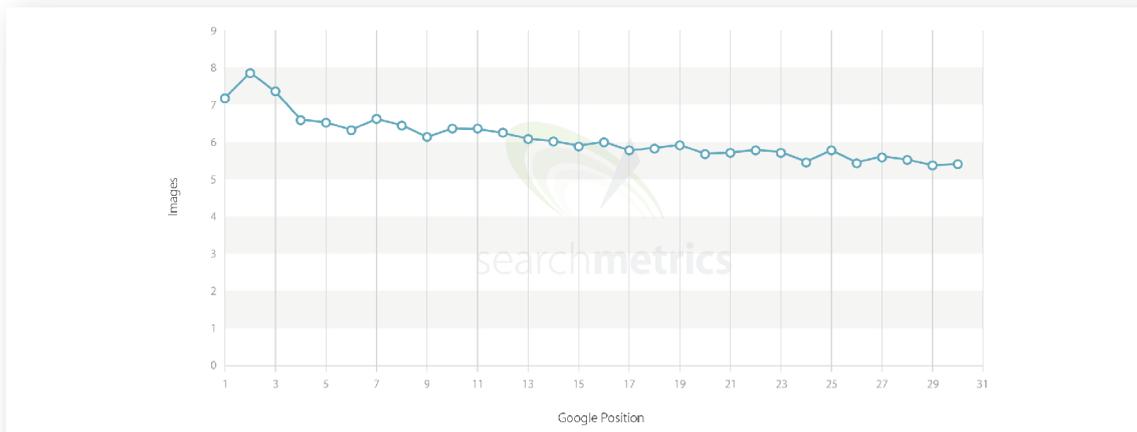


Figure 30: Average ranking - number of image files

Even if one ignores smaller image elements, the data shows that the more pictures there are per page on average, the better the page is ranked.

2.4 Internal Linking: Housekeeping is the beginning and the end

The internal link structure of a domain is a vital component of domain performance. To distribute link juice¹² optimally, factors such as „number of links“ and „link text“ are crucial. Not only are the number of links that point to a ranking URL relevant, but the link structure on the URL itself is crucial.

Shown below is the number of internal content links compared on the sites analyzed:

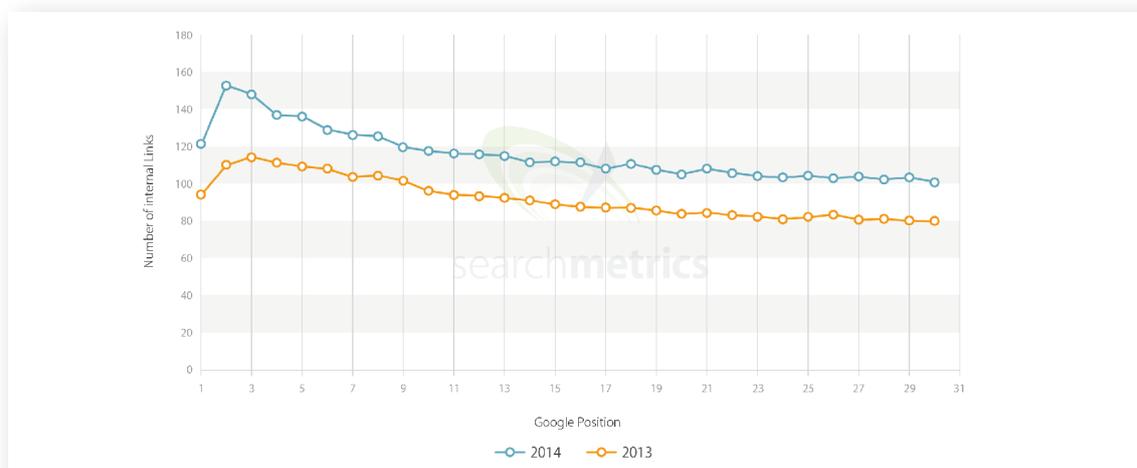


Figure 31: Average ranking by number of internal links 2013/2014

With both the correlation and the average values rising this year, you might think it would be true that more links = better ranking. However, this is not the case.

This feature is probably even more capped than the corresponding Content features. Regarding internal links, it is not correct to simply say that „More is better,“ it depends very heavily on optimal site architecture.

¹² Every link has a value. This value can be optimally distributed according to the principle of inheritance using a good internal link structure according to the hierarchical structure of a domain. The term link juice is meant metaphorically here.

The increase may be partly explained by the growing amount of content, for example. The hypothesis could be that more content = more links generally, because the absolute number of external links has grown year over year.

It is clear that pages in the top search positions have comparatively more internal links than pages further down the rankings. It can also be seen that Brand pages in the top rankings have fewer links on average than the pages below them.

Note that in every case this is the average figure. The ideal position seems to be about 120 to 140 internal links per URL, but this is on a case-by-case basis with no scale. A link should always make sense to both the page and to the expectations of the user within the structure of the page. Broken, irrelevant, and unnecessary links should be removed as a housekeeping priority. Many companies and website operators make the mistake of focusing on external links rather than trying to optimize the user experience and internal link flow.

The average number of links is increased on pages with a robust menu structure and internal links throughout the page. It's not only the number of internal links, but internally linked keywords that also play a role in the optimal distribution of link juice.

We asked ourselves an interesting question in our last analysis: Is it a good idea to internally link a ranking keyword on a well-ranked page to another page on the site? If so, this would imply that the page that gets the link would be as relevant to the keyword as the source page for the link itself. If this were the case, would the higher ranked page pass its link juice to the lower ranked page, and would it then lose its own ranking, sooner or later?

The implication of this would be to use the reverse-principle, and link well performing pages to themselves. This question provides the basis for the analytical approach in the following graph:

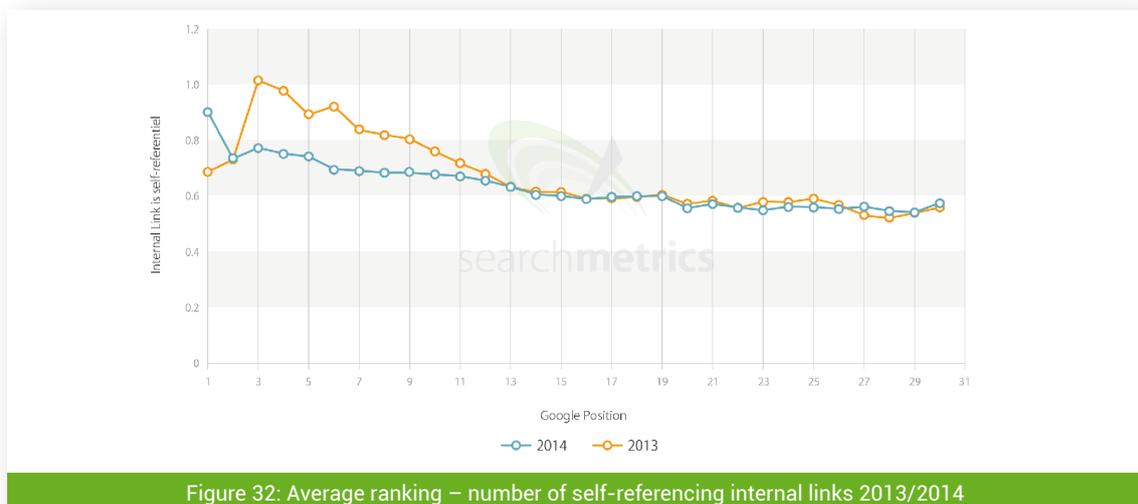


Figure 32: Average ranking – number of self-referencing internal links 2013/2014

The average for the top result increased slightly in 2014, while the rest of the top ten listings on SERP 1 were below those for last year.

The Y-axis of this graph shows the average number of self-referential internal links, the X-axis shows the Google rank.

9/10 sites at position 1 in SERPS have at least one internal self-referencing link – in 2013 it was 7/10. Values for the following items fell, however.

Results from our study show that sites with self-referencing keywords on pages (including many pages related to 'Shopping') usually linked from a site-wide menu or via a breadcrumb link back to itself.

Important Note: For a keyword to improve its ranking, internal links to another page should be avoided at all costs. Many pages that climb the rankings link the keyword internally to itself on a technical structure basis.

2.5 Adlinks: In total less advertising – except for Brands

In 2012, there were significant negative correlations with good rankings in terms of advertising integrations, even for the Google product AdSense. Two years ago, well-ranked pages placed less advertising than poorer-ranked pages. These correlations were slightly more positive in 2013, before continuing the original trend this year.

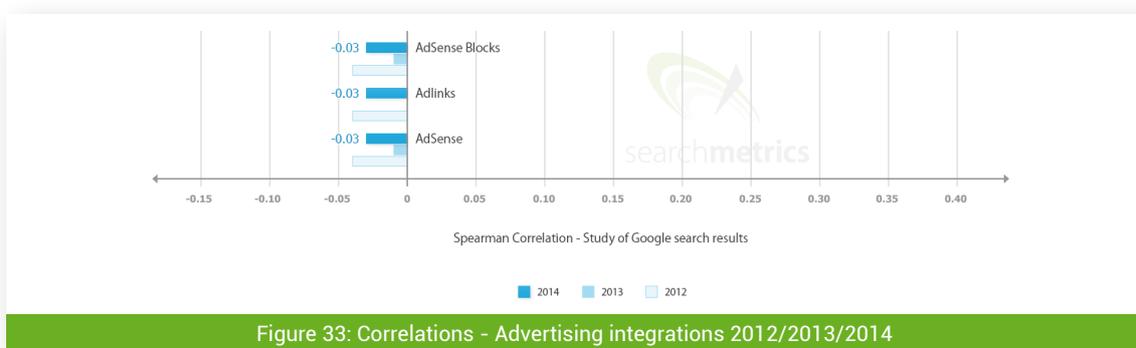


Figure 33: Correlations - Advertising integrations 2012/2013/2014

This graph shows that the existence and the number of advertising integrations decreased on average. The following stand-alone graphs illustrate this in detail.

The Y-axis of the graphs below are scaled from 0 to 1 and then multiplied by a factor of 100 to obtain the percentage of pages at the respective position, shown on the Y-axis.

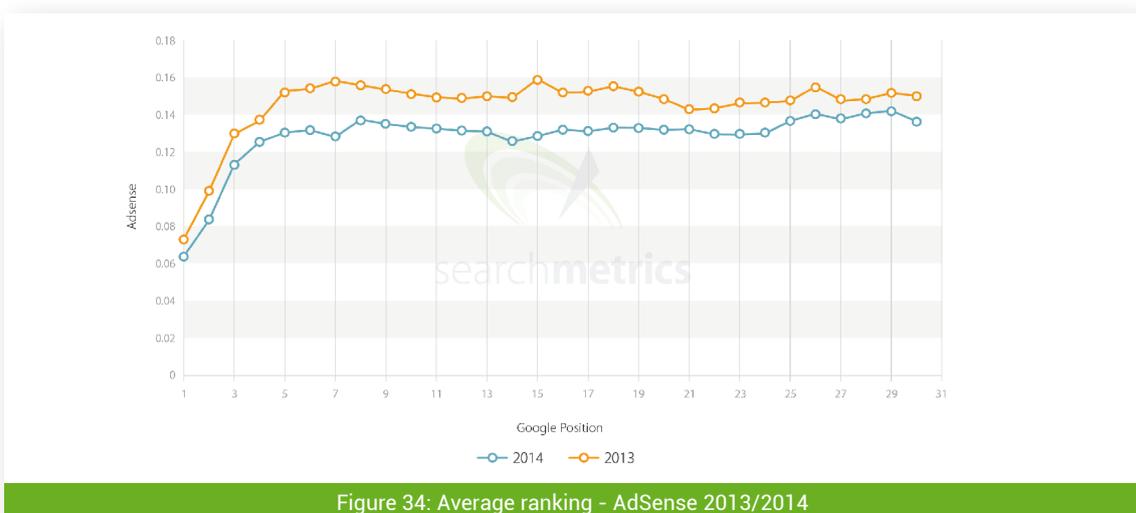


Figure 34: Average ranking - AdSense 2013/2014

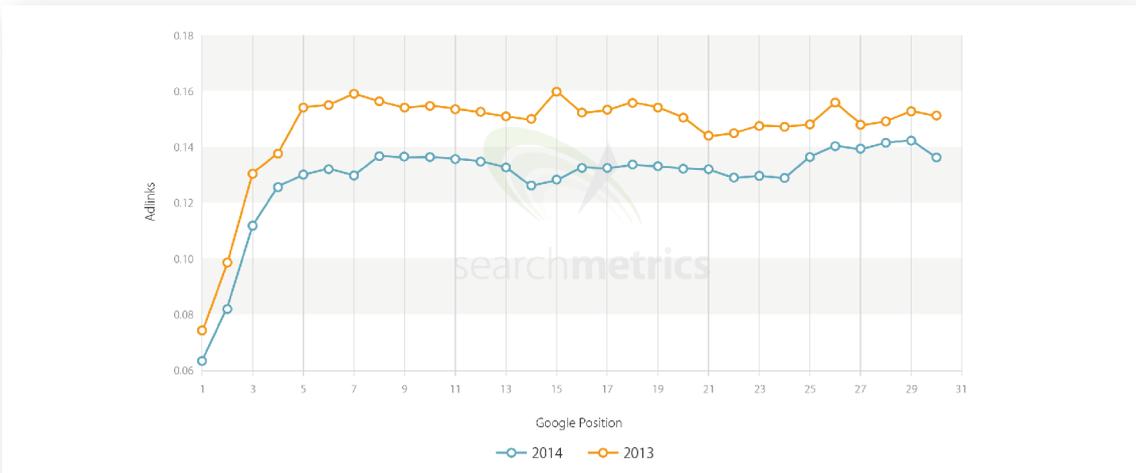


Figure 35: Average ranking - Adlinks total (including AdSense) 2013/2014

The primary message from these two charts is that, especially for URLs that have reached the top 3 search result positions, there are on average fewer pages with at least one integrated advertising feature, (AdSense or another form of Adlinks) than URLs in the lower positions.

AdSense is included on most pages as another form of Adlink. Most of the pages in the top 30, consequently, use the Google advertising integration itself.

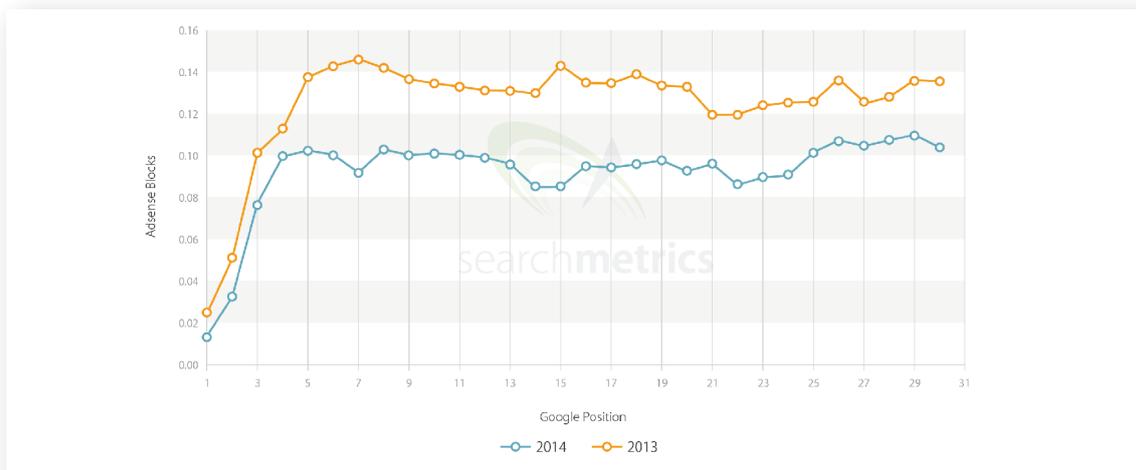


Figure 36: Average ranking - number of AdSense integrations 2013/2014

Based on the number of AdSense integrations on high-ranking sites, the top 3 ranking URLs have fewer advertising integrations than those sites in the second half of page one and the top half of page 2 in SERPs, which have about the same number. There is a slight rise in integrations from position 24.

It is an interesting fact that the number of AdSense integrations for sites at position 1 has not changed very much, but there has been a considerable reduction in advertising integrations on pages climbing the rankings in the top 30 positions.

Author Integrations In Google SERPs

Last year, the correlation between Authorship and rankings was examined, and we found that almost no correlation existed. Google has recently told authors to no longer use photos in their profile. The Circles notification has also disappeared, leaving only the name. There is one exception, Google News. Here the author image stays, but appears much smaller and its position has been moved (see screenshot).



Figure 37: Google SERP Snippet before adjustment - Organic (screenshot)¹³



Figure 38: SERP snippets after adjustment. Above: Organic, Below: News¹⁴

On 06/28/2014 the time had come, the author pictures were removed. Here's the first page of search results for the keyword „Panda 4“ in a direct before-and-after comparison:

Some studies show that author photos have a positive impact on CTR. Why Google removed them remains an open question. Searchmetrics will take this development as an opportunity to create our own analysis of how this adaptation of SERPs impacts CTR and we will compare results with and without an author photo.

¹³ Screenshot: 06/23/2014

¹⁴ Source: <https://support.google.com/webmasters/answer/1408986?hl=en>

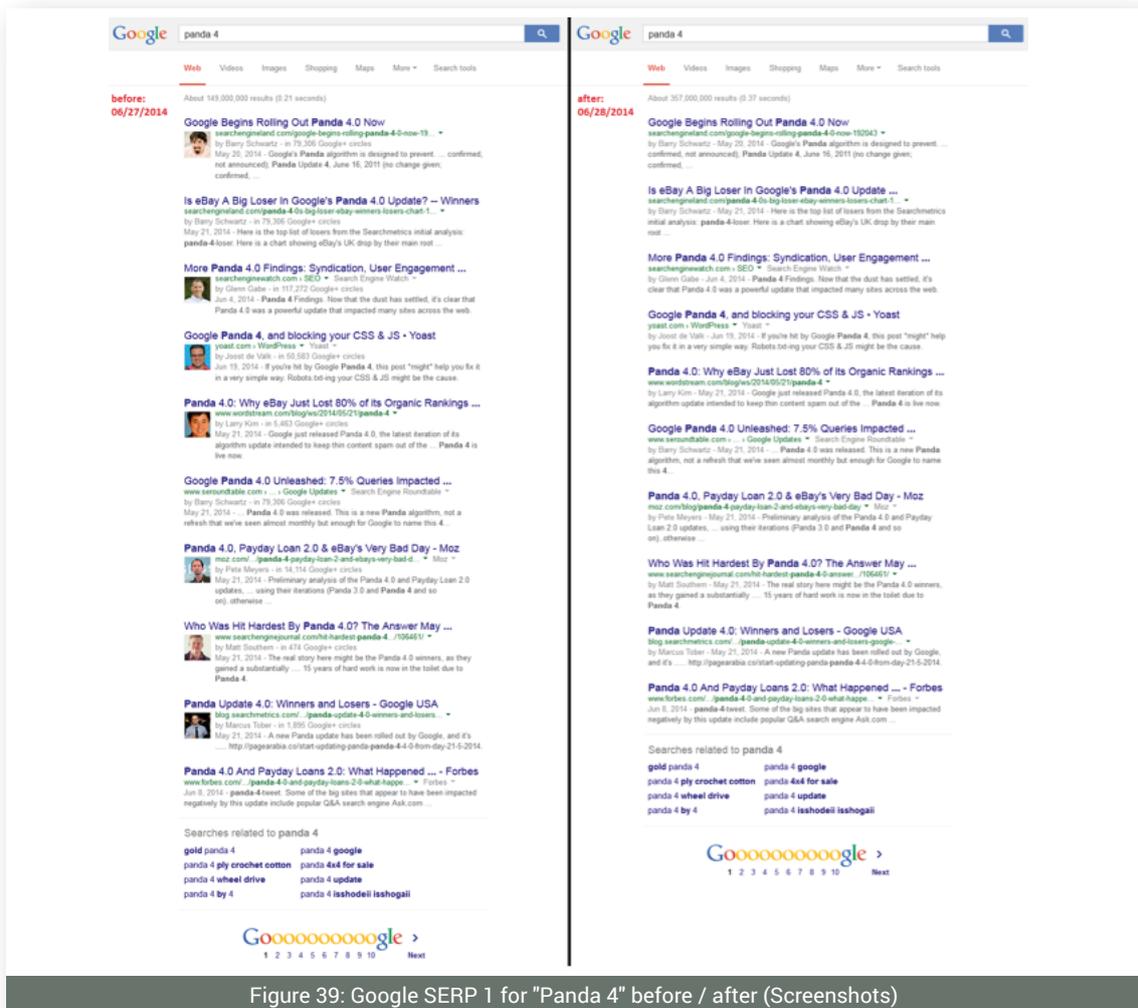


Figure 39: Google SERP 1 for "Panda 4" before / after (Screenshots)

Summary

Over the last few years, content quality has become more and more important as search engine algorithms become more and more advanced. High-ranking URLs have more text and more media integrations than in either 2012 or 2013.

A good internal link structure is an important quality factor. Housekeeping and a good site information structure are crucial too.

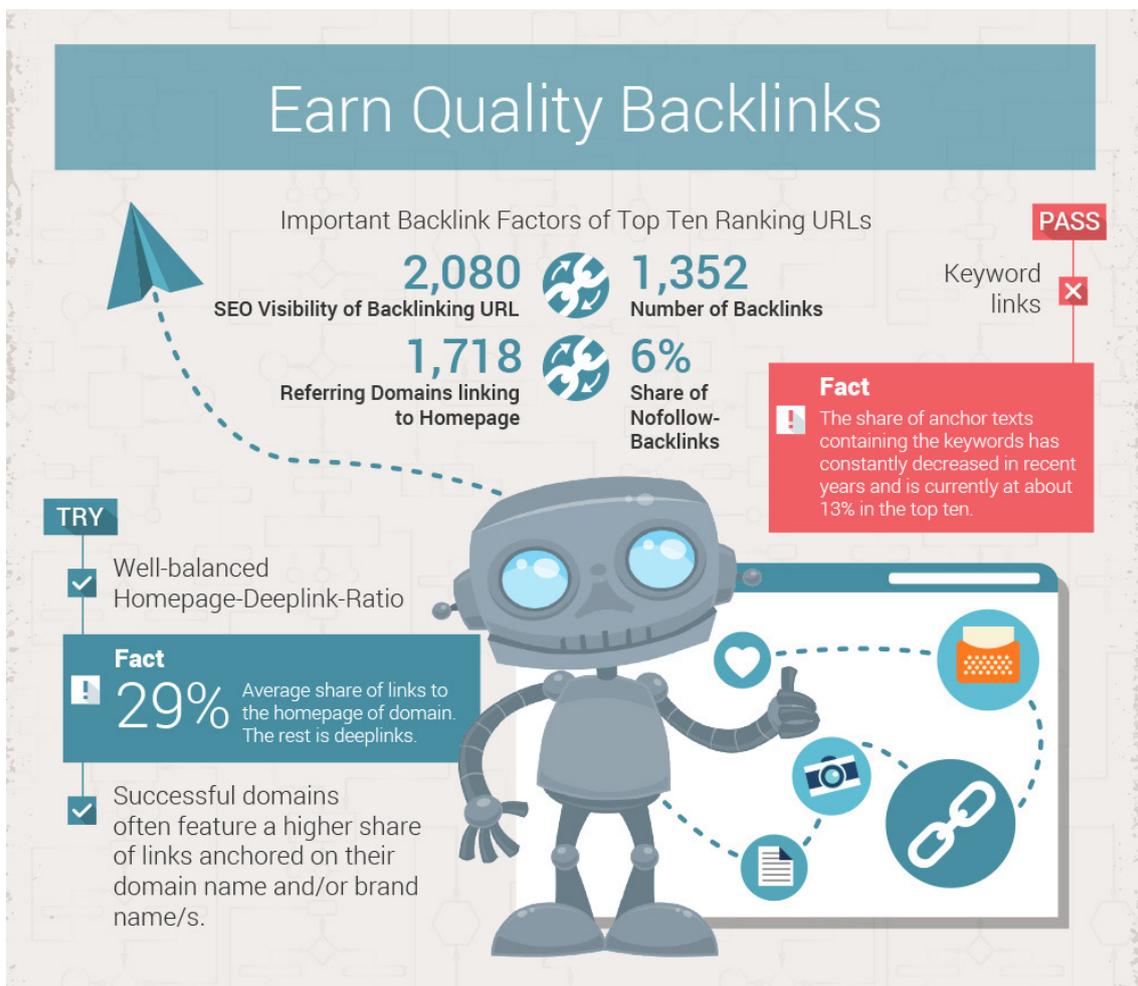
Advertising integrations have decreased for URLs in the top 30 - less so for Brands, more for URLs from position 5 and after.

The absolute primary focus for content, however, is quality. To improve ranking for content, it is no longer good enough to concentrate on the number of links or keywords on the page itself.

Optimization is, increasingly, a holistic discipline. It's not enough to rank for one relevant keyword – content must now be relevant to the topic and include several related terms. This became very clear in our Proof and Relevant-Term analysis, and in the Searchmetrics study on the impact of Google's Hummingbird update to the diversity of landing pages and domains¹⁵. As a result of this update, it became apparent that search result diversity decreased and holistic content ranked better for more keyword queries.

¹⁵ cf. searchmetrics.com/en/knowledge-base/hummingbird/*

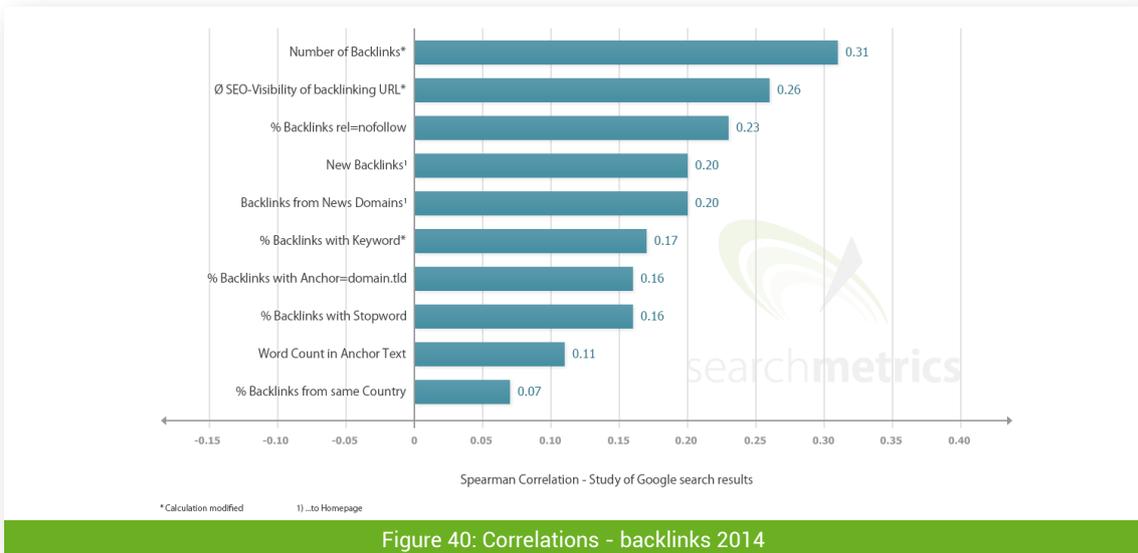
At a glance: Infographic – Backlinks



3. BACKLINKS

Backlink quality factors are probably the most important SEO metric on the off page side. Google's success is strongly based on „PageRank“ which has always been oriented towards backlinks, and remains so to this day. Links from other sites belong to search engines and form a detailed picture of the evidence of relevance based upon the subject matter linked to. Opportunities to manipulate the system are constantly being minimized with targeted algorithm changes from search engine companies.

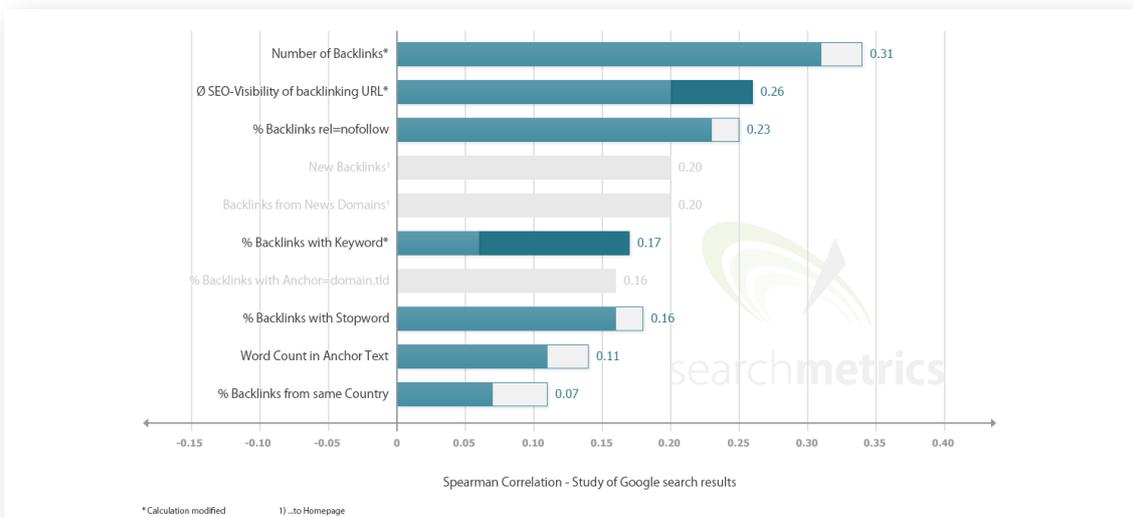
Consequently it's not only the number of links, but also the quality of the backlink structure of a page, itself subject to various factors, which is crucial to a good ranking. In order to understand the diversity and complexity of the backlink cluster calculation, the pool of factors used to analyze backlinks was extended in this report – primarily in an attempt to analyze the influence of „Brand Links“.



Since backlink metrics are among the oldest metrics used, and the most studied, we weren't expecting to see big changes from the previous year with regard to differences in the top 30 positions.

Backlinks: Development Compared To Last Year

The following graph shows the development of the primary factors from previous years.



Graphic 41: Correlations - changes to backlinks 2014 / 2013

The lower correlation for the number of backlinks is immediately noticeable. This had fallen from 2012 – 2013, however, link quantity still has a high correlation with good ranking. An examination of how the other correlations have developed will form part of the following sub-chapters.

3.1 Link Source: The higher the quality, the better

The source of any given link is highly important and relies on the following factors shown in the graph below:

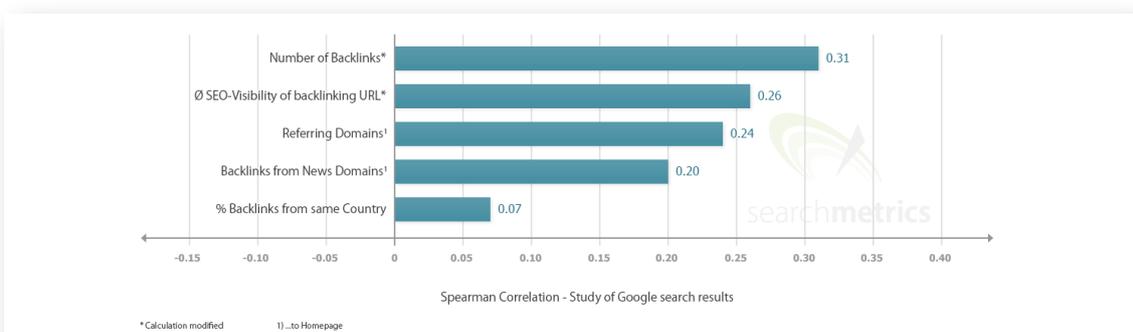


Figure 42: Backlink Correlations - Link Source

The Number Of Backlinks Remains Important

The graph below shows the marked difference between pages at position 1 and lower-ranked URLs in terms of the number of backlinks present:

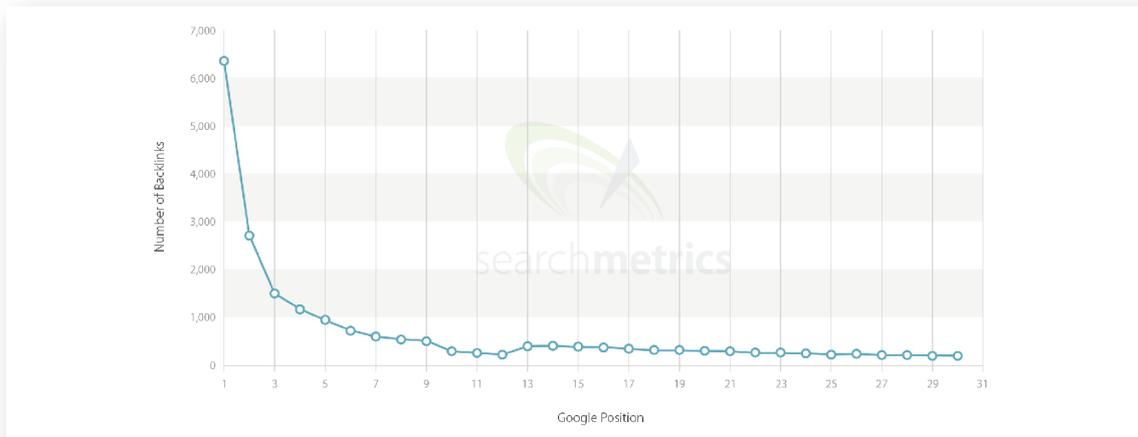


Figure 43: Average ranking - number of backlinks

This graph shows the number of links in relation to the ranking of the page. On average, pages ranked 1st have over 6,000 backlinks – with sites in second place having fewer than half this number, and sites at position 30 having just under 167 on average.

Comparing these figures to 2013, it is noticeable that the number of backlinks from sites in the top positions has fallen significantly, which explains the lower correlation. This is due to improvements in Searchmetrics crawling, with the whole backlink area being subject to certain dynamics such as links constantly being updated and deleted from sites. The backlink profiles examined were clean and the data more accurate than in the previous year, where re-crawling non-existent links was not noted. This means that the data from 2013 is not broad enough to compare with 2014, and that the absolute number of backlinks in 2013 might have been higher.

Better Ranked URLs Have Much Better Links

„SEO visibility,“ also known simply as “Visibility”, has its origins in SEO analysis software tools. It is calculated from a blend of factors concerning the keyword rankings of a keyword pool, and is equivalent to a traffic measurement (not measurable by outside parties). It acts as a basic indication of the quality of a Web page for comparative purposes. Searchmetrics takes into account search volume, information regarding keyword characteristics (navigational or informational search) and the layout of the search result, as well as keyword rankings. In short, the higher the SEO Visibility of a page, the more likely it is to generate organic traffic from search engines¹⁶.

A look at the correlation of the “SEO Visibility of the linking URL” factor reveals that the relevance of the traffic coming from a link from ‘URL B’ to ‘URL A’ seems to be crucial to the ranking of URL A.

In short, backlinks from sites with high SEO visibility seem to be much more valuable than links from sites with low SEO Visibility.

¹⁶ The SEO Visibility of domains is calculated for many countries and is available for free at <http://suite.searchmetrics.com/us/research>

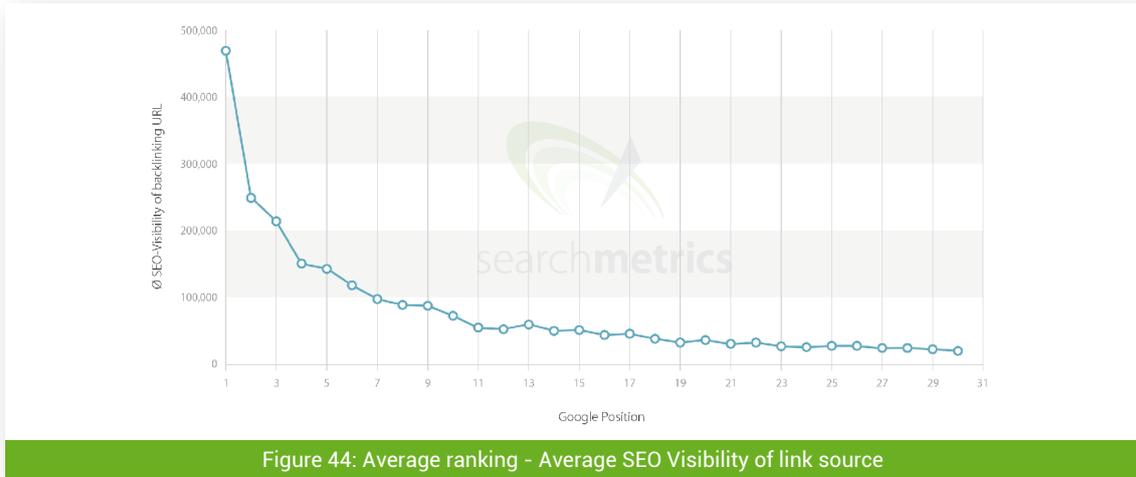


Figure 44: Average ranking - Average SEO Visibility of link source

This year the feature was calculated differently. In 2013, 100 backlinks from the URL were analyzed at random, whereas in 2014, 100 backlinks were analyzed for each landing page, using those with the highest SEO Visibility as the basis of the calculation. Now the analysis shows the highest-ranking links in prime position.

High-ranking URLs are boosted by their backlink profile compared to those sites lower down the rankings. Pages at position 1 in SERPs have links from sites with a higher SEO Visibility score than those sites lower in the rankings.

Number Of Referring Domains To The Homepage

Many of the following features share similar characteristics and trends at the top SERP positions. This has been reinterpreted and redefined for this analysis, and relevant definitions are detailed in the chapter on 'Brand Factor.' It should be noted that bigger brands are separate from smaller ones. Major brands are responsible for very high values starting at position three, whereas smaller brands have often surprisingly low values at position one. Wikipedia results are calculated using averages, which explains why values at position two (where Wikipedia is often found) are often lower.

The first of the new features shows the correlation between the number of Referring Domains to the homepage of the examined URL:

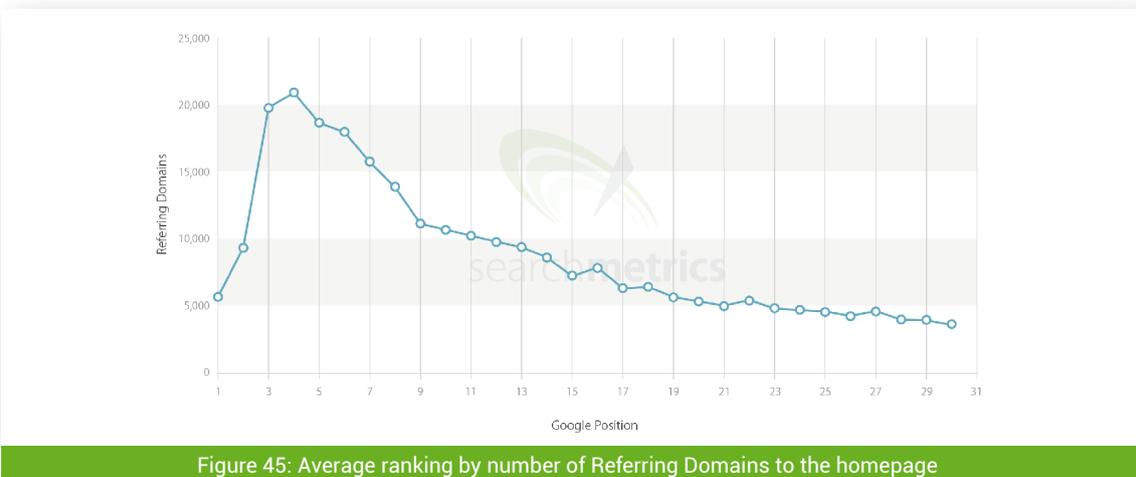
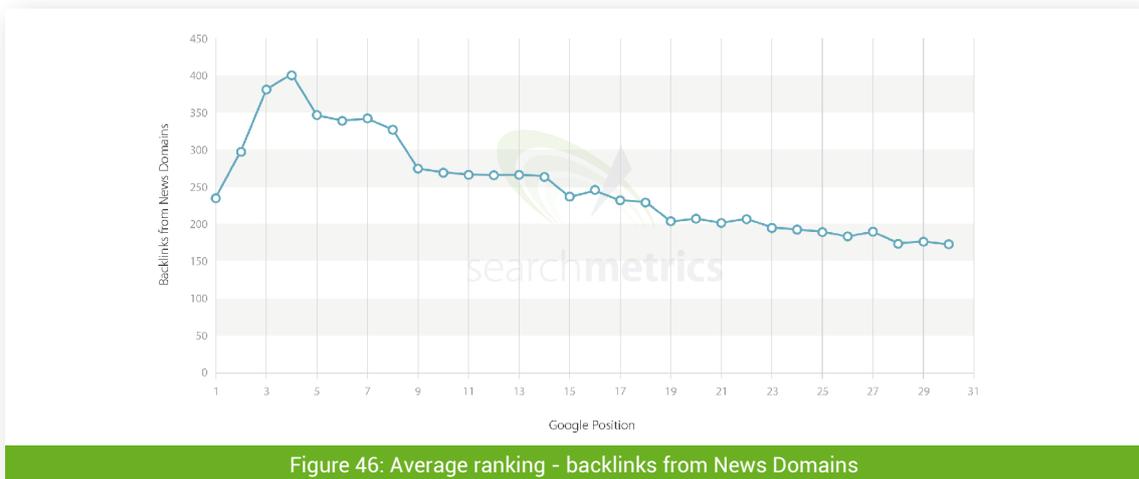


Figure 45: Average ranking by number of Referring Domains to the homepage

As you can see, bigger Brands have many more Referring Domains than the rest of the Top 30, which are almost in the same range.

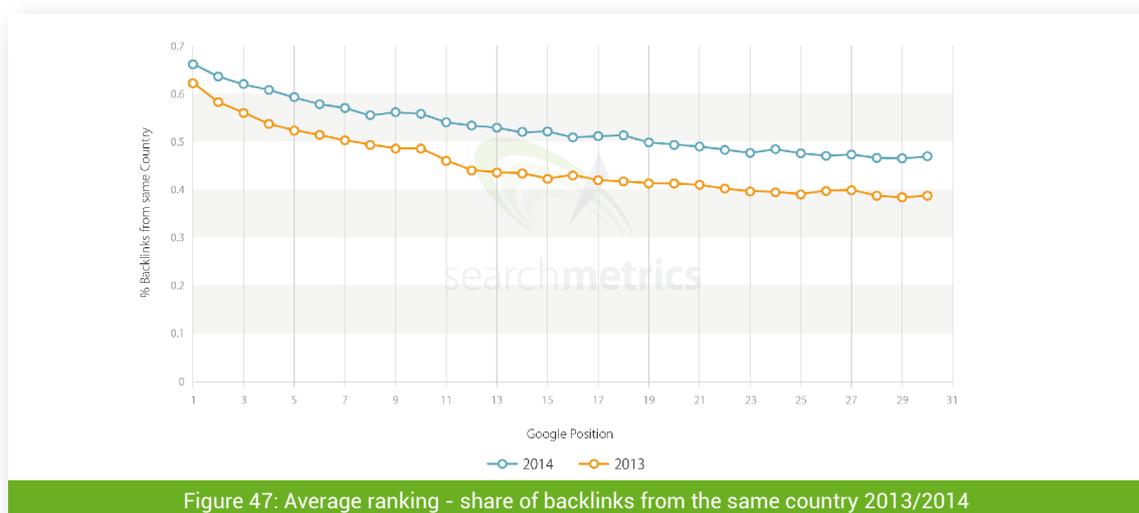
Number Of News-Domain Links To The Homepage

Larger brands have significantly more links from news domains than any other URLs. Bigger Brands often generate media coverage, and this is reflected in this metric:



Number Of Links From The Same Country

There has not been much change to the proportion of backlinks from the same country. The correlation is slightly lower, but remains very similar to the curve from the previous year.



URLs on the first page of search results have, on average, a higher proportion of links from other countries.

3.2 Link Quality: The need for quality continues to rise

The quality of backlinks remains crucial. The following graph shows this clearly:

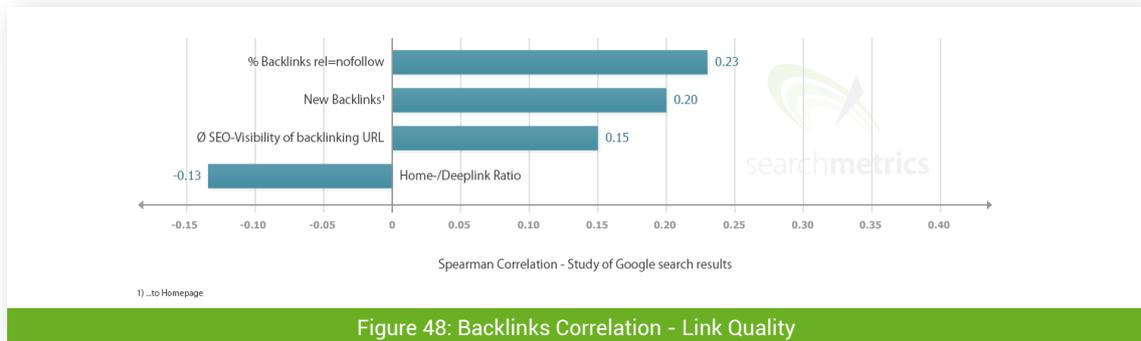


Figure 48: Backlinks Correlation - Link Quality

New Backlinks

The following new correlation is similar to other new features. Here, big brands have significantly more new Links. However, the curve is uneven. Interestingly, smaller brands in the first search position have, on average, fewer new links compared to URLs from SERP position 3 onwards.

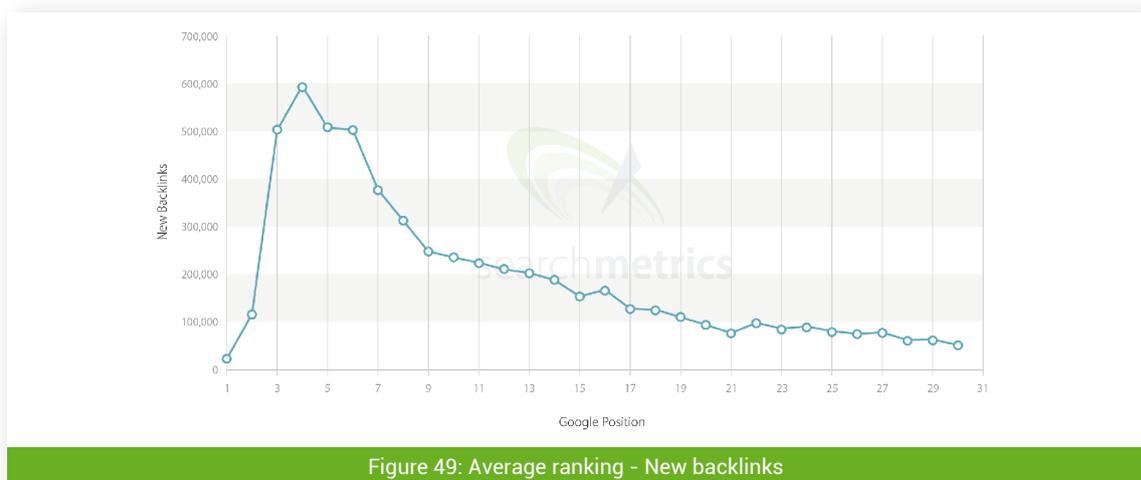


Figure 49: Average ranking - New backlinks

Share of Nofollow Links

„Soft“ factors at a Backlink level seemed to be more relevant last year. Good ranking pages have a significantly higher proportion of multi-word anchor texts, and no-follow links in their backlink profile, shown in the following diagram (Y-axis: value multiplied by a factor of 100 results in %):

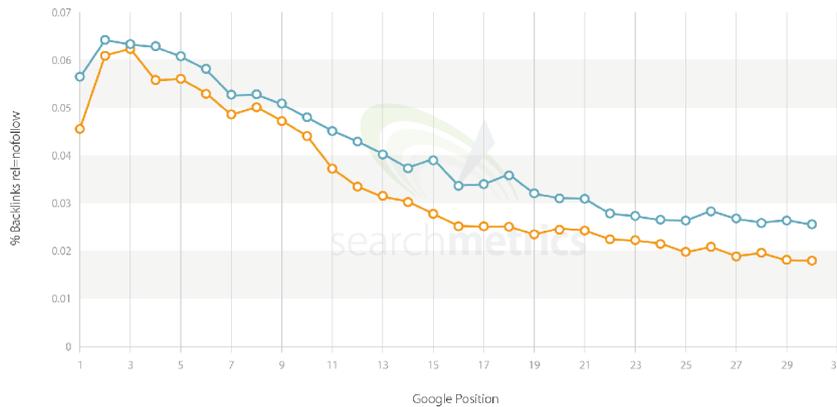


Figure 50: Average ranking - share of no-follow backlinks 2013/2014

The curve for 2014 is very similar to that of the previous year – but the correlation has decreased.

Average Link Age

Examining the average link age results in a continuously decreasing curve with the y-axis describing the age of the links in days. Better ranked URLs have, on average, older links. This is an indication that better-ranked pages in SERPs have been in existence longer. This correlation is related to the correlation of the number of backlinks. Those sites with more links also have more old links.



Figure 51: Average ranking – Link age

Homepage / Deep Link Distribution

The following graph shows the average proportion of links to the homepage of the domain compared to Deep Links. There is a slight negative correlation, with a rise at positions 1 and 2, which is attributable to niche brands.

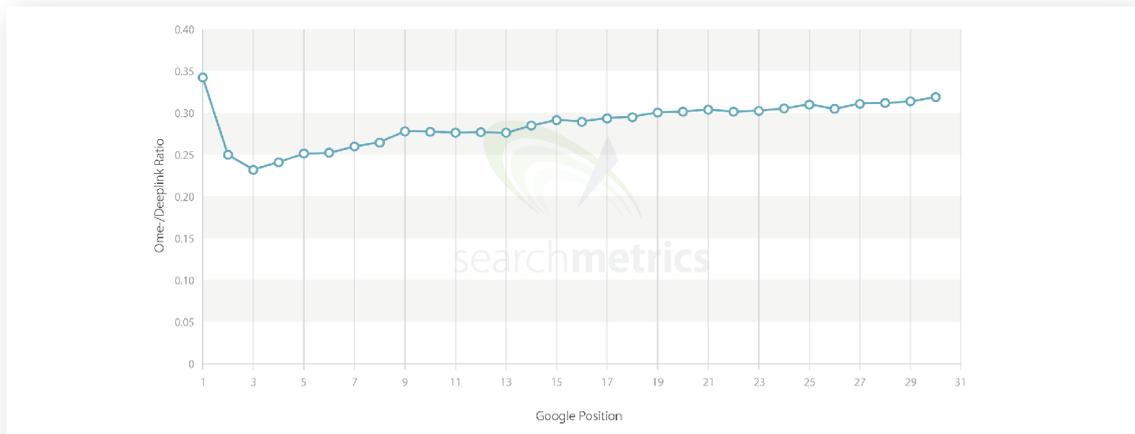


Figure 52: Average ranking - links to homepage

With the exception of the two top search result positions, better ranking sites have more deep links. The average is just under 29%.

3.3 Link Anchor – Link Diversity Becomes More Relevant

The third link area represents the nature of link anchors - the text surrounded by the HTML link attribute - on the link-giving site. In the past this was a heavily optimized factor.

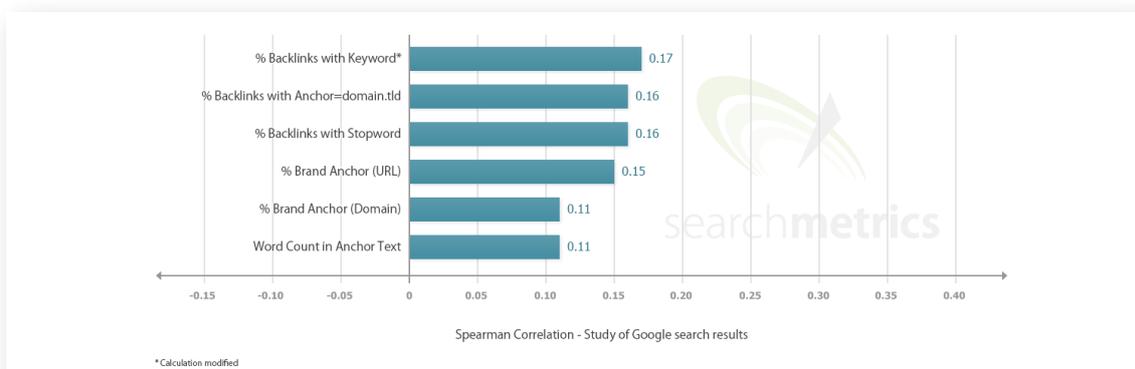


Figure 53: Backlink Correlations - Link Anchor Text

Keyword Anchor Texts Are Rare

The days of „hard“ keyword links were numbered last year. 2012 showed a positive correlation for this factor for the Top 30 positions, but this decreased significantly during 2013, and was adjudged to be just barely positive. In 2014, we again calculated a positive correlation for this feature. Has the number of links with keywords on the top search result positions increased again? No, on the contrary, there is another explanation

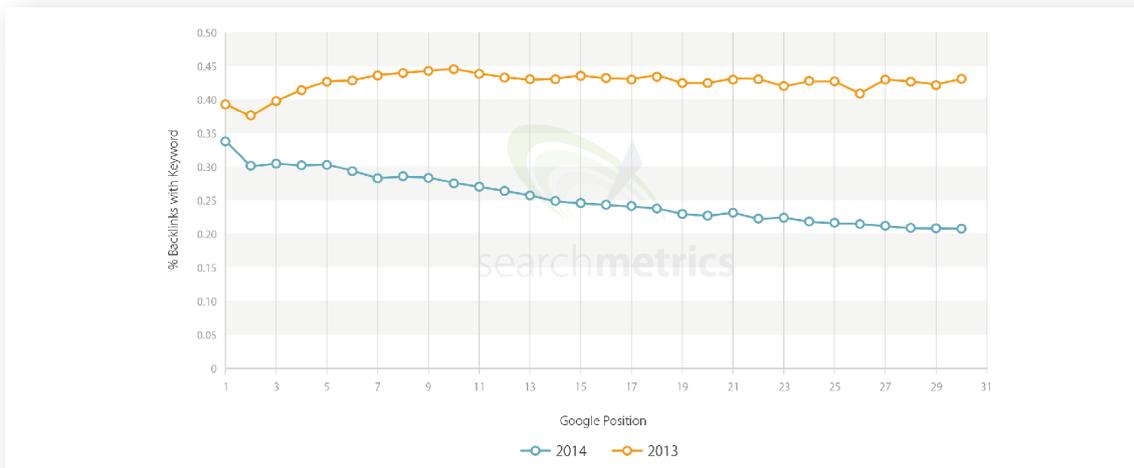


Figure 54: Average ranking - share of backlinks with keyword 2013/2014

This is a good example of correlation vs. averages. Last year, a slightly positive (near-zero) correlation was present. This year the feature correlates much more positively with good rankings.

A look at the average curve can shed some light on this. The ratio has fallen by more than a third over the year. The reason for this positive correlation is that the Brand factor is more pronounced and the curve is now slightly downwards.

At positions below 5, this is likely to have happened within the last few months as the result of a revised approach to Anchor text optimization, and may have had something to do with link degradation. There is now much less keyword-heavy link building, and the Penguin update and its numerous iterations seem to have hit their mark.

The curve for this factor looks a lot more natural, with the peak at position 1 due to brand links.

Backlinks with a keyword in the backlink still have their place and rank well (the value for the Y-axis by is multiplied by 100 for the percentage), however, pages in the top 30 positions now show only 25 percent of backlinks with the keyword in the anchor text.

Backlinks With Stop Words

For the „keyword links“ factor, the inclusion of Stop Words in the link text has become more natural than in the previous year. The link no longer includes a single keyword, it is common for webmasters to link several juxtaposed words.

This factor correlated positively with good rankings in 2013, and not much has changed in 2014:

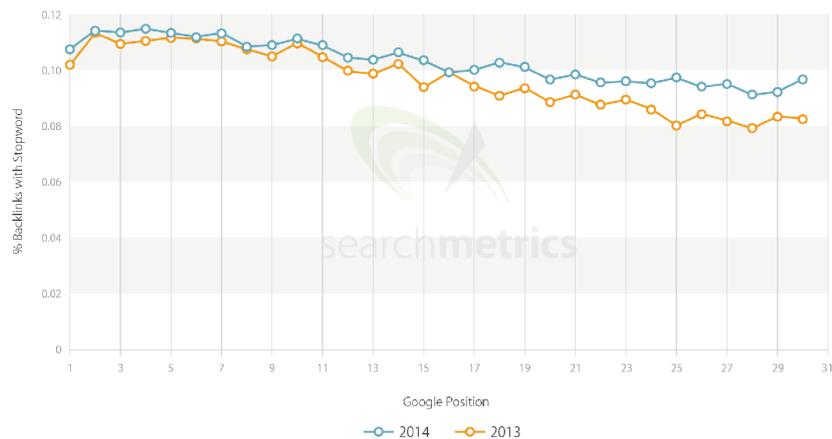


Figure 55: Average ranking - Ratio of backlinks with Stop Words 2013/2014

Even if the correlation value has dropped slightly, the curve is almost identical to that from the previous year. Google seems to have found a balance.

It seems that Google treats it as „unnatural“ when individual URLs are always linked with the same keyword or anchor text.

It seems highly unlikely that a large number of webmasters (authors, etc.) would link independently of each other to the same page using the same link text. Google has decided to treat this – probably correctly – as deliberate manipulation of a link profile where hard keyword links exceed a certain degree of tolerance, and it has subsequently devalued this factor to an extent.

Number Of Words In Anchor Text

In order to attain the highest possible rankings, an increasingly sophisticated link building strategy needs to recognize positive trends in link quality factors. These positive correlations were more pronounced in 2014 than in 2013.

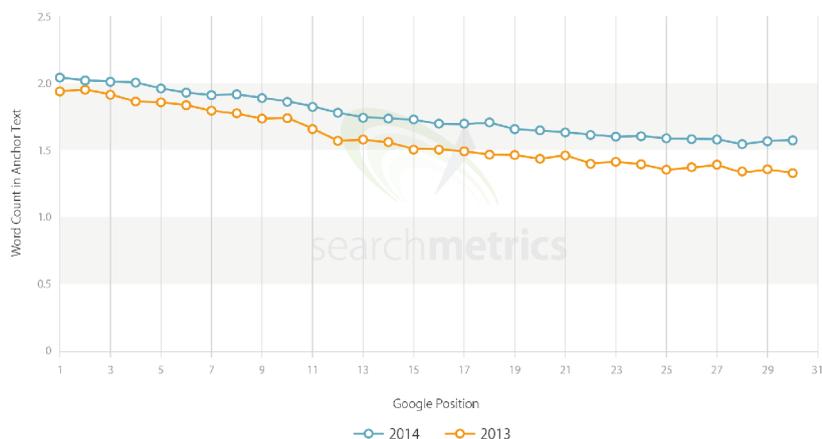


Figure 56: Average ranking - number of words in the anchor text 2013/2014

Again, the curves for 2013 and 2014 are almost identical. This means that the linking behavior of webmasters seems not to have changed much in the year.

In addition, the number of words in the anchor text - although minimally increased compared to last year – seems to have been capped. No one uses entire paragraphs as links, for example, let alone multiple Webmasters, so that the curve would rise.

Nevertheless, it remains that higher-ranked sites still have more links with more words in the Anchor.

Proportion Of Backlinks With Anchor Text = Domain.Tld¹⁷

This feature is explained further in the chapter on the Brand factor, but since the following factors are strongly Link or Anchor Text-heavy, they form part of this chapter too.

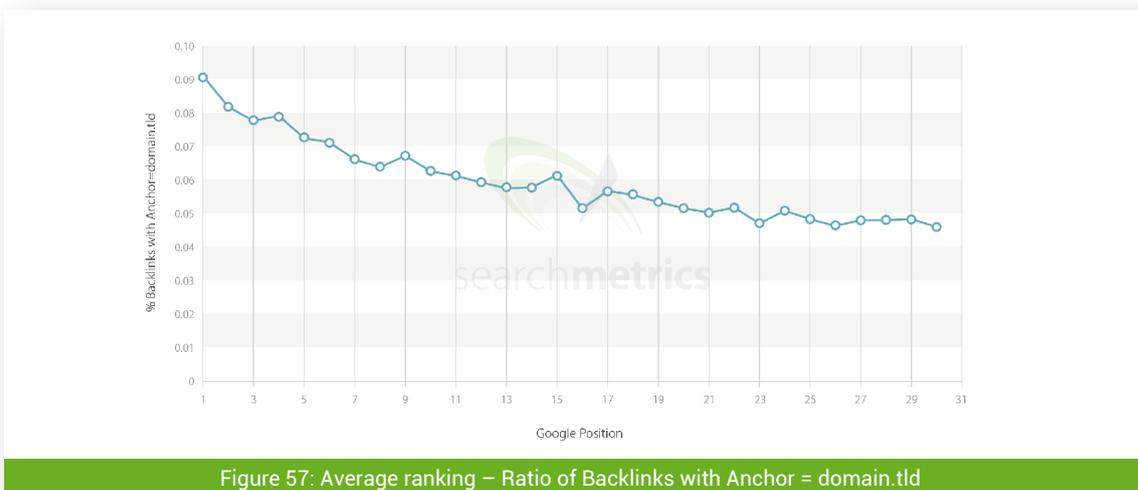


Figure 57: Average ranking – Ratio of Backlinks with Anchor = domain.tld

Shown here is that portion of the Backlink Anchor Text of a URL + Domain TLD (Top Level Domain) (i.e. „time.com“ or „people.com“). SERP position 1 ranked URLs have more Deep Links containing the domain Anchor. This has strong brand identity, because the corresponding Webmaster for the entity (the domain name) connects through anchor text of its own, surrounded by a deep link attribute. In principle, the content that lies behind this deep link is promoting this content.

Proportion Of Brand Anchor Texts To Domain And URL

Based on this, the proportion of Brand Anchor texts to other link texts were analyzed for all anchor texts of the URL or domain. The Brand Anchor is that anchor text which occurs most often - excluding the domain.tld anchor text¹⁸. An example of such a Brand Anchor would be „people.“

¹⁷ TLD = Top Level Domain / domain extensions like .com, .org etc.

¹⁸ Specifically the brand name was supposed to be extracted, not the domain name (with domain extension), which is yet often linked more frequently as anchor text.

Anchor texts (20 of 85,300)

Anchor text	Referring Domains	Frequency	In %
people	6,871	177,545	3.10
people.com	3,129	29,873	0.52
people magazine	2,847	23,134	0.40
People	2,698	512,096	8.94
n/a	1,410	70,955	1.24
People.com	1,068	7,296	0.13
source	1,021	6,556	0.11
here	917	4,409	0.08
People Magazine	759	9,194	0.16
www.people.com	501	2,581	0.05
[img]	471	11,366	0.20

Figure 58: Search Metrics Suite - Link texts of people.com (screenshot)

This screenshot from the Searchmetrics Suite shows that the domain URL www.people.com is most commonly linked with the anchor „people.“ For this analysis, the counting was adjusted so that the anchor domain.tld was excluded.

Based on this, Brand Anchor ratios were calculated first for all Anchor texts of the respective URL, and secondly for all anchor texts on the entire domain. Interestingly, the two correlations show, a different picture.

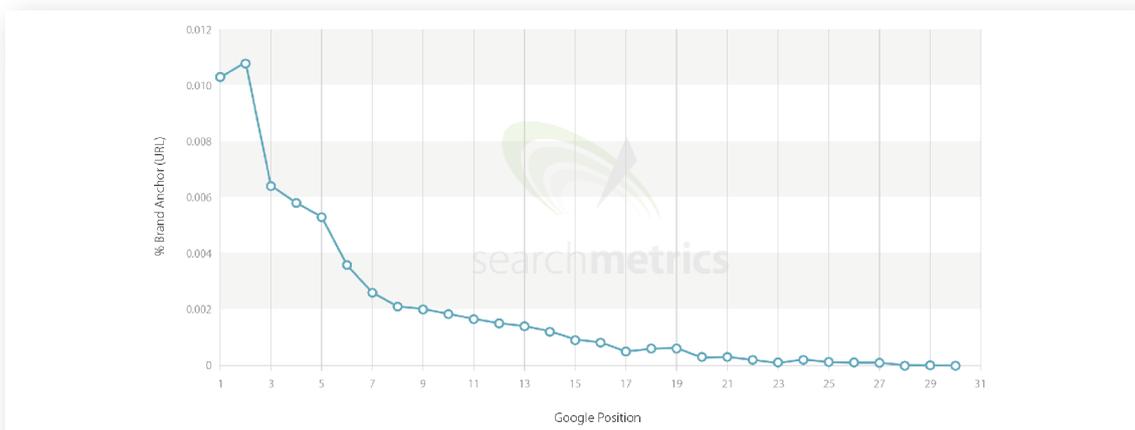


Figure 59: Average ranking - Brand Anchor ratio to URL

First, let's look at the proportion of links with Brand Anchor texts in the URL. Note that the scale on the y-axis is very detailed, and the curve falls fast with the exception of the first ranked position. Deep End URLs are often the link with the brand-name domain as anchor. The proportion is only about 1.1 percent, however, even in URLs at the second search result position.

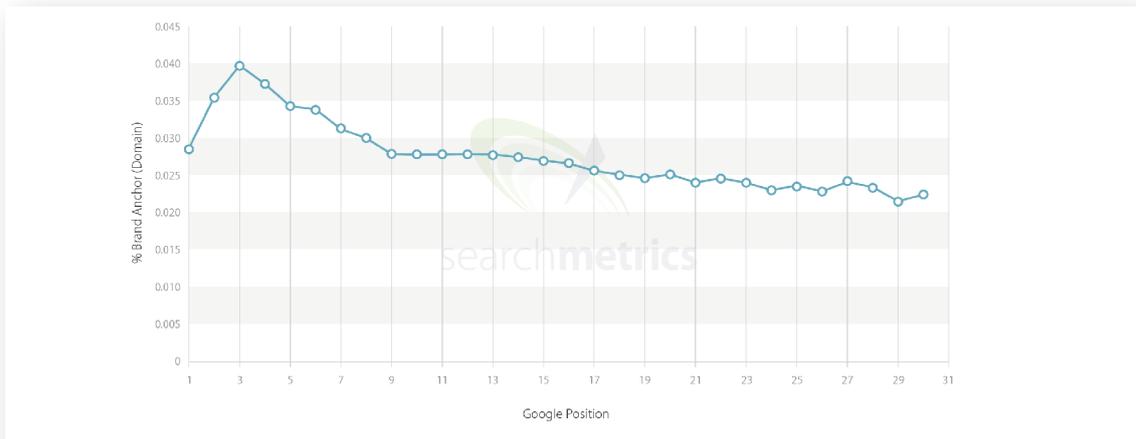


Figure 60: Average ranking - Brand Anchor ratio on domain

This proportion is slightly higher with respect to the entire domain, but the curve here is distinctly different. First, the correlation is linear, but falls off again, and secondly, the first two search result positions are due to the Brand factor. Here again, small brands, and the removal of Wikipedia results are the cause. URLs on position 3 have the most backlinks with only the brand as the anchor for the domain.

Here, too, however, the differences are very small. The maximum distance between the curve points is only slightly more than 1.5 percent. This could be because domains are often linked with the domain ending in the Anchor.

Summary

A good backlink profile is still a cornerstone and an important SEO metric. Last years' study showed that not only the sheer number, but increasingly the quality of the links, play a role.

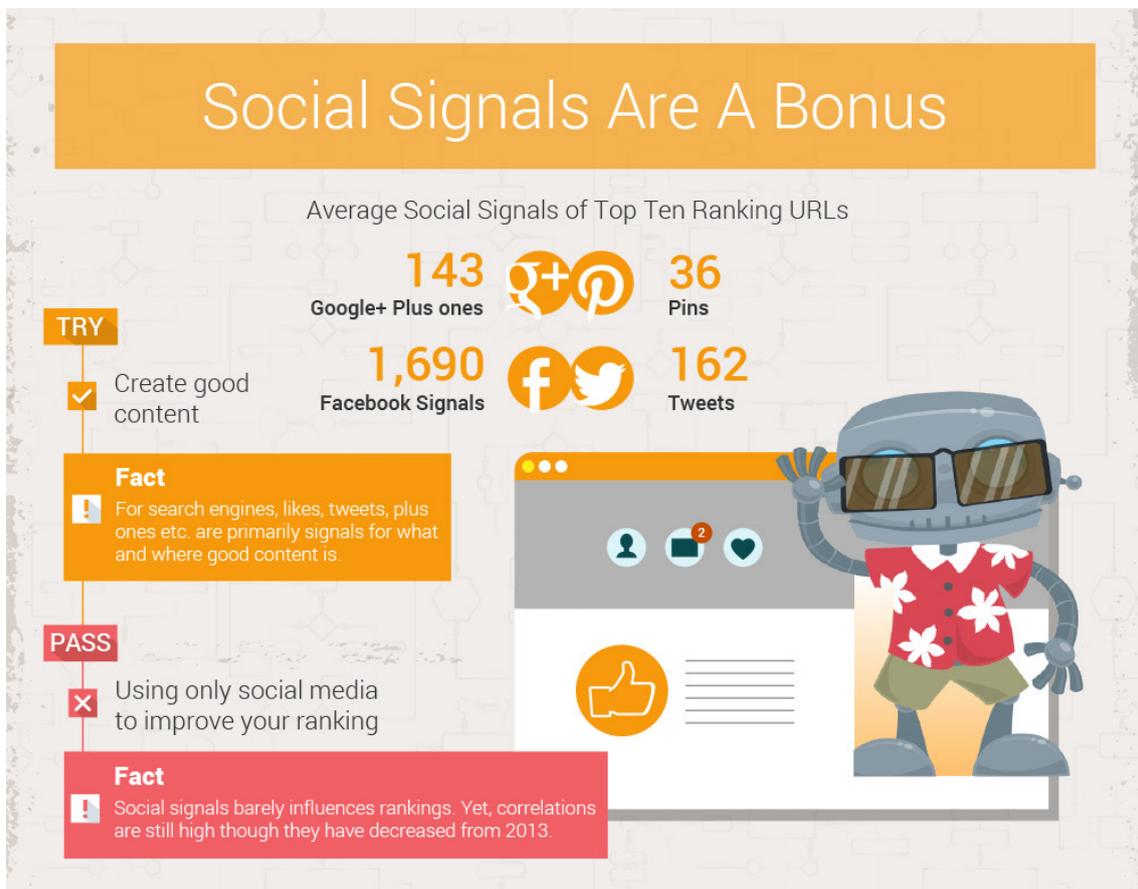
In the top ranked positions, URLs not only have significantly more links than any other site, but also have a link profile that is in the broadest sense „balanced,“ or natural. This is characterized on average by a higher proportion of backlinks with higher SEO Visibility, with Stop Words, from the same country and / or nofollow attribute and more words in the anchor text.

Keyword factors have undergone significant devaluation by Google via the Penguin Updates - particularly during 2013. The expression of the proportion of links that contain the keyword in the anchor text has become more natural again this year. The correlation has indeed grown, but the proportion of all links has been reduced by about one third on average. This does not include brand links. URLs in the top positions have more links that contain the „domain.tld“ string.

Larger brands - and in general pages that have climbed to SERP position 1 - also have more new links, more Referring Domains, more links from news domains and more deep links.

With regard to the backlink profile, a balanced mix seems to be the secret of good rankings. Moreover, the quality of the link source plays an important role. Google's algorithms, and their own Quality Rater can unmask link farms much better these days. The backlink profile is now a blend of a lot of layers of quality-driven factors. SEOs who are looking for single, isolated signals that can be unilaterally scaled and manipulated, will just not find them anymore.

At a glance: Infographic – Social



4. SOCIAL SIGNALS

The field of social signals is still the area with the highest correlations on average, but it is also the area with the most misinterpreted factors. At this point it is important to underline that correlations should not be confused with causal relationships¹⁹.

Objective fact remains removed from causality, however pages in Google's top-ranked positions generally have a very large number of social signals. In other words, good ranking URLs have many shares, Likes, Comments, +1s and tweets.

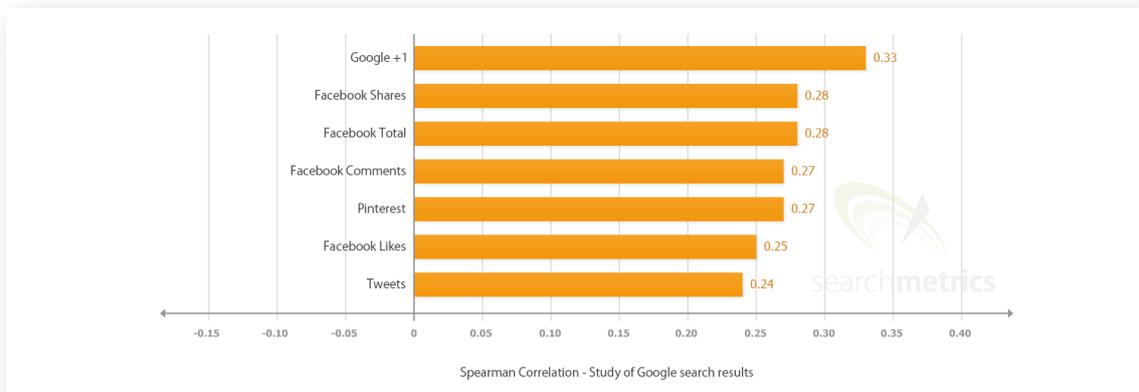


Figure 61: Correlations - Social signals 2014

Sites that have risen in the SERPs have strong positive correlations to social signals. The detailed graphics in the following sub-chapters show this clearly. Before we dive into that, let's look at the development of ranking correlations with social signals over the year..

Social Signals: Developments In The Previous Year

In the following graph you can see that correlations fell for the first time. As early as 2012, social signals were strongly correlated with good rankings, and in 2013 they were more frequent and more pronounced features of well-positioned URLs in the SERPs. Now, the values have decreased for the first time over all networks and individual factors.

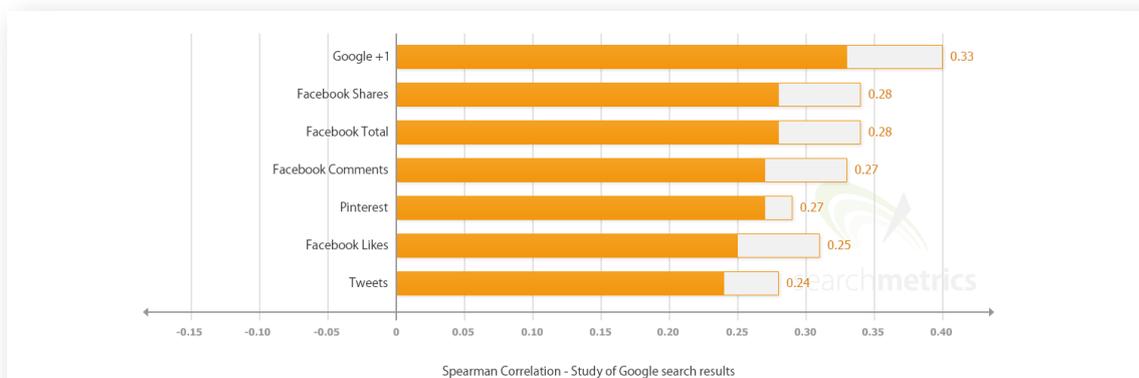


Figure 62: Correlations - Changes to Social signals 2014/2013

¹⁹ See 'Causality ≠ Correlation' in the introduction

Because every social signal factor has increased, the differences between the first and last search positions have been reduced.

Social Signal: Correlations Decreasing, But Still High

The correlation of social metrics with better rankings remains high, with recommendations and related signals from the social graph strongly related to good Google rankings.

The detailed graphics for individual networks show very pronounced curve ratios from which to calculate strong correlations, and little has changed in the interpretation of the data from the previous year.

The following graph and comparison to 2013 includes new, aggregated data that was not included in the previous report. It should be pointed out here that the 2013 values are not the values from last year's study, but values from the mid-year, based on more data becoming available. This makes the figures much more directly comparable with those for 2014, and therefore, more accurate.

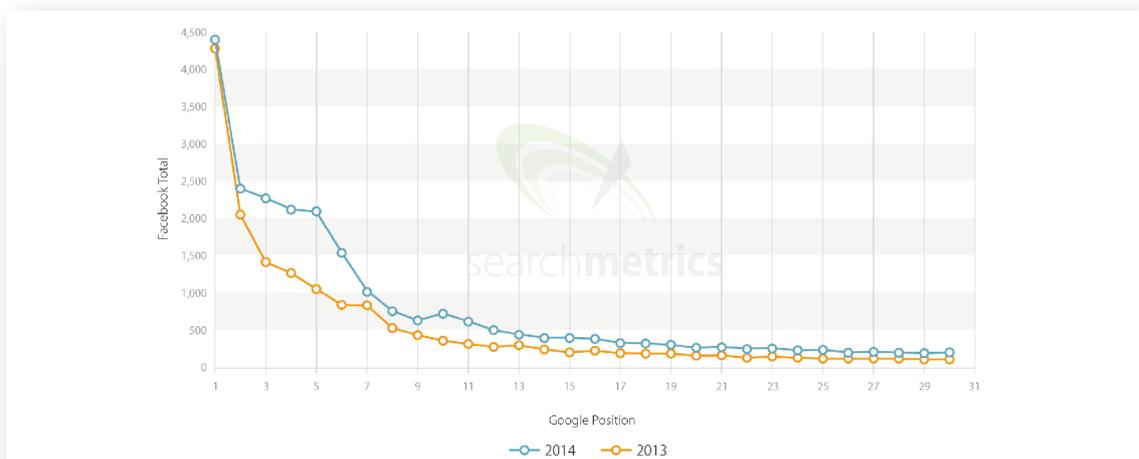


Figure 63: Average ranking - Facebook signals 2013/2014

The „Facebook Total“ factor is shown here for all Facebook metrics (Shares, Likes and Comments) and is a summarized correlation of those factors. Interestingly, the number of Facebook signals has increased in positions 3-6. This supports the Brand analysis, which has been adopted for this study.

The average number of tweets from the Twitter network - an information aggregator rather than a Social Network - per ranking URL in proportion to the Google position, however, is well below the figures for Facebook signals. While the average number of tweets for sites in the top position has decreased in comparison with last year's figures, the correlation with rankings is still clearly visible.

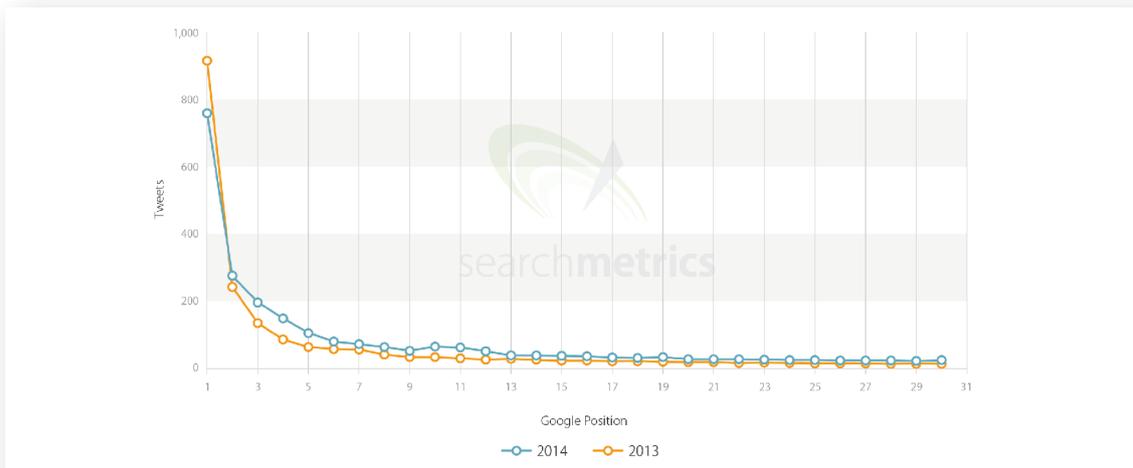


Figure 64: Average ranking - Twitter signals 2013/2014

Pins - signals from Pinterest, a platform that allows the sharing of content primarily in the form of images via bulletin boards - are relatively sparse as a social signal for page and ranking, but even here the distribution reflects a fairly good correlation.

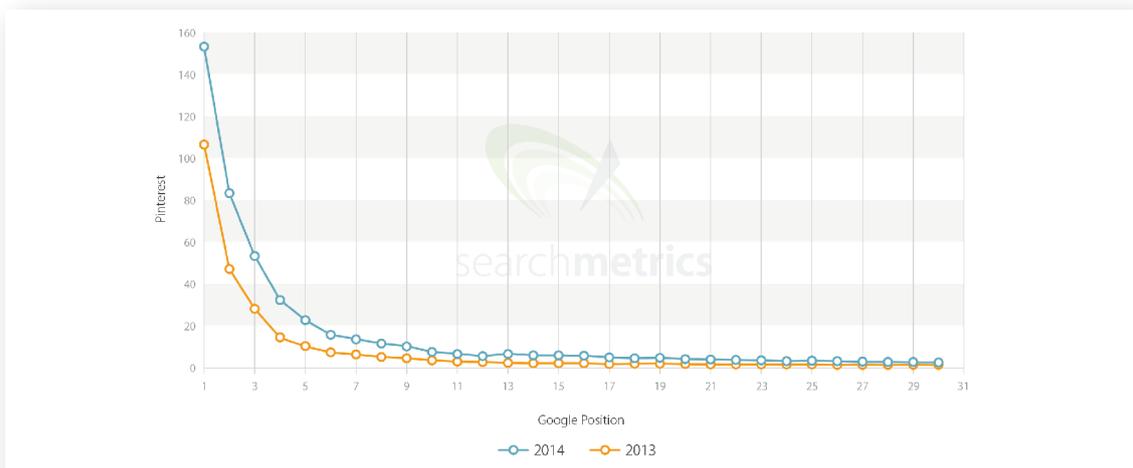


Figure 65: Average ranking - Pinterest signals 2013/2014

In summary it can be stated that social media factors include the following:

The graph curves have a profile characterized by a strong decrease in values from position 1 to 2, and then a leveling of the curve on the X-axis. URLs in first position in SERPs have the most social signals by far.

The slight anomaly of values at position 2, where the metrics for pins, tweets and Facebook signals are lower than for those at position 3 on average, and which appear as a slight 'blip' in an otherwise smooth curve, seems to be due to the large number of Wikipedia results which are typically seen at position 2. First position from Google's point of view seems to be reserved for Brand URLs.

The results show that the top-ranked brand has many more social signals than Wikipedia in position 2. URLs in position 3 have risen, as these receive many more social signals on average than the related Wikipedia result.

The analytical approach for Google+ is the same as for the other social networks, but with one small difference, namely, the values for these Google network signals for sites at position 1 have notably increased.

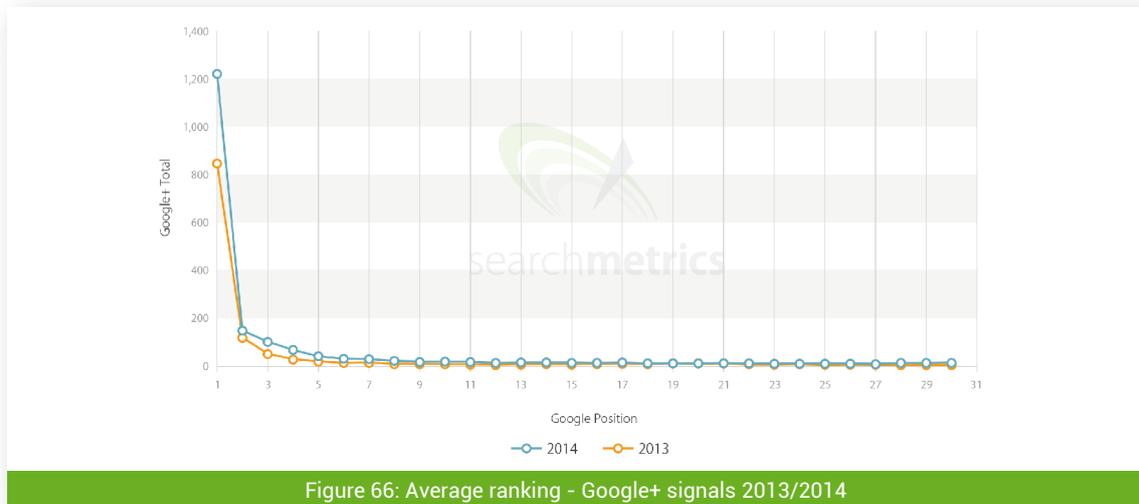


Figure 66: Average ranking - Google+ signals 2013/2014

The fact that the correlation has decreased, suggests that URLs in lower positions have more +1s on average this year.

Summary

Social signals are factors that correlate strongly to better rankings. In terms of correlation values and in terms of absolute averages Facebook is ahead of Google, followed by Twitter and Pinterest.

The absolute values have risen moderately across all these social features, but correlations across the top 30 positions have increased. As the total number of social signals continues to rise, there are now smaller differences between search result positions.

The question of how social signals directly affect rankings remains. As noted in our analysis, higher-ranked URLs have more social cues than those sites further down the ranks, but Google has emphasized that they are not using social signals as a ranking factor. At the very least, a high number of social cues related to a site will indicate new content being added.

Last but not least, social signals definitely play a role in direct traffic, brand awareness, and the overall online performance of a domain. This is understandable from a search engine perspective, in that good content is much more common on social networks, and search engines want to provide good, relevant content.

5. THE BRAND FACTOR AND WIKIPEDIA

The Brand Factor is a phenomenon we have mentioned previously. It runs through this entire study, and it strongly influences results for nearly all correlations shown on the graphs. Brands have risen to the top positions of the SERPs, even if there are certain factors they do not fulfill which lower-ranked sites do.

However, there are gradations in the definition of brands, which add an interpretative element to this study. Three different kinds of Brand factors can be analyzed.

Three Different Kinds Of Brands

The Brand Factor has been mentioned over the last few years, and in this study we cover it in more detail than before. Some of the features and their corresponding correlations are much easier to interpret when the Brand Factor is taken into account.

The starting point was the observation that Brands rank higher. Since there are fewer brands than domains, let alone URLs, we would expect to find that a higher position in SERPs would have fewer domains. The results show that this is the case – apart from at position 1.

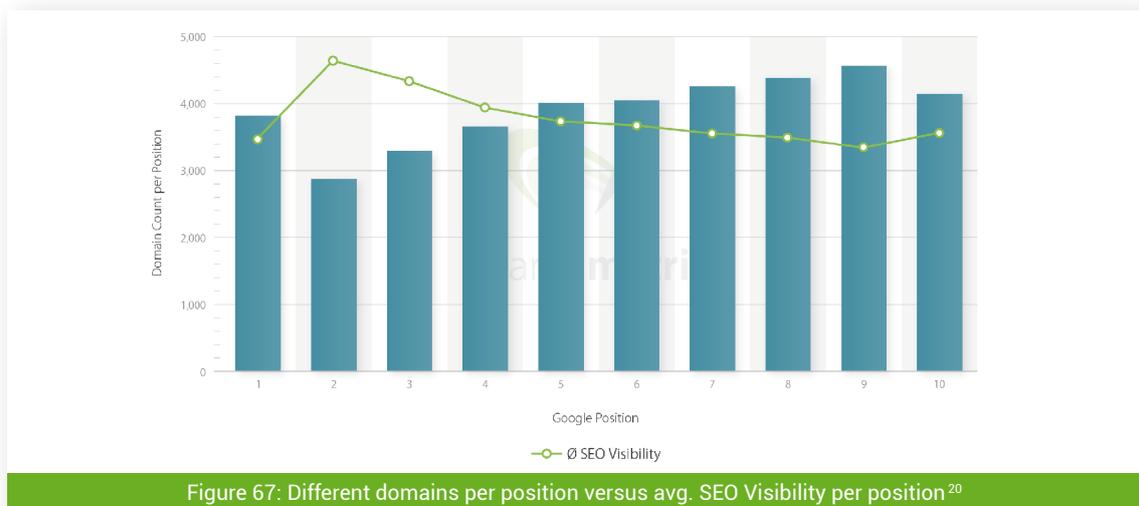


Figure 67: Different domains per position versus avg. SEO Visibility per position ²⁰

The analysis shows that brands rank better, with the number of domains steadily decreasing from positions 10 to 2, with only position 1 showing different results. More domains are shown at position 1, slightly more than at position 3.

In contrast, the average SEO Visibility for all keywords is also shown in the above chart. There is an anomaly here: The curve is nearly the opposite of the domain count per position, with the highest SEO Visibility being noted where there are the fewest different domains.

If we drill down even further into the analysis to examine larger domains in the top ten ranked positions, the following picture emerges:

²⁰Y-Axis scale (SEO Visibility) disregarded



Figure 68: Number of Wikipedia rankings per position

The analysis for Wikipedia is highlighted because the values of the other domains in the graph are not as significant. The curve in this case is very similar to that for the „Wikipedia URL“ factor.

As you can see, Wikipedia often ranks at position 1, but is more often 2nd. From position 3, the curve falls significantly.

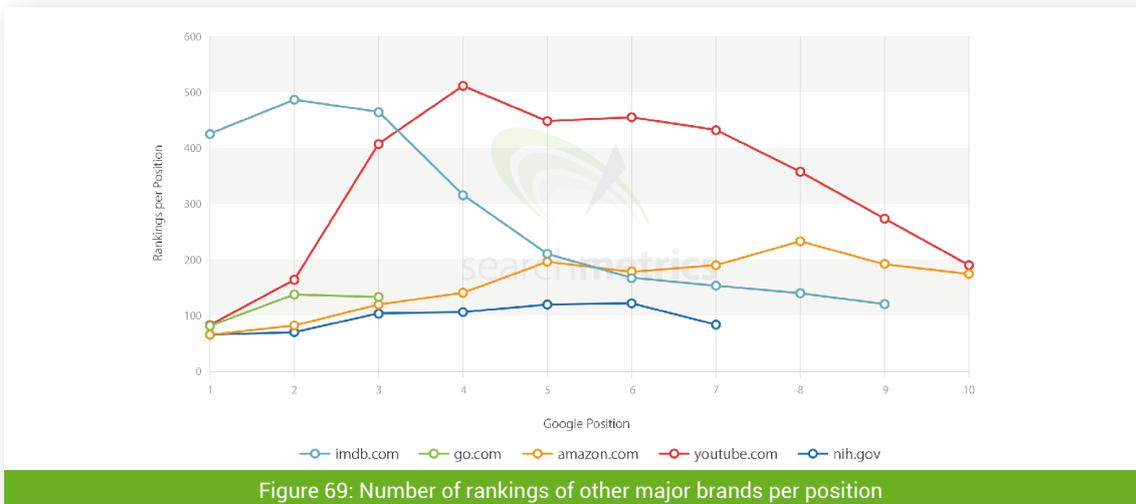


Figure 69: Number of rankings of other major brands per position

The ranking data for Wikipedia and several other brands was based on an analysis of the 10 best rankings per position, over all examined keywords. Go.com, for example, isn't in the rankings for these after position 3.

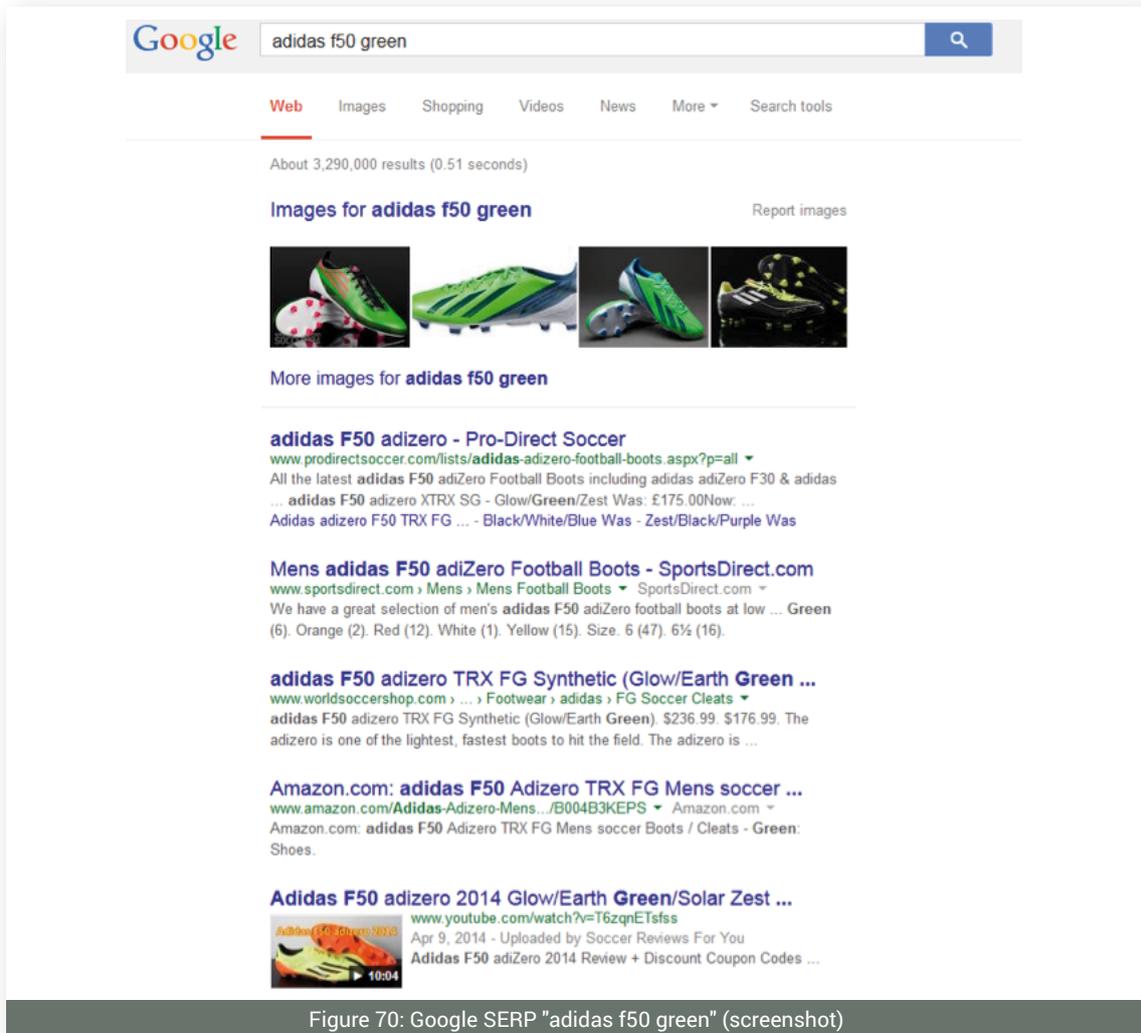
It seems that almost all major Brands rank at position 2 rather than position 1. As brand rankings differ, though, it should be noted that larger Brands rank better for many of their keywords, but Google seems to give great influence to another factor, niche Brands.

The Searchmetrics Brand Analysis

1 Niche Brands

Niche brands usually have smaller domains with specific content relevant to their niche. They very often rank for niche terms on position 1 - the more specific the better. For more generic, yet topic related keywords, they often rank in the top 10.

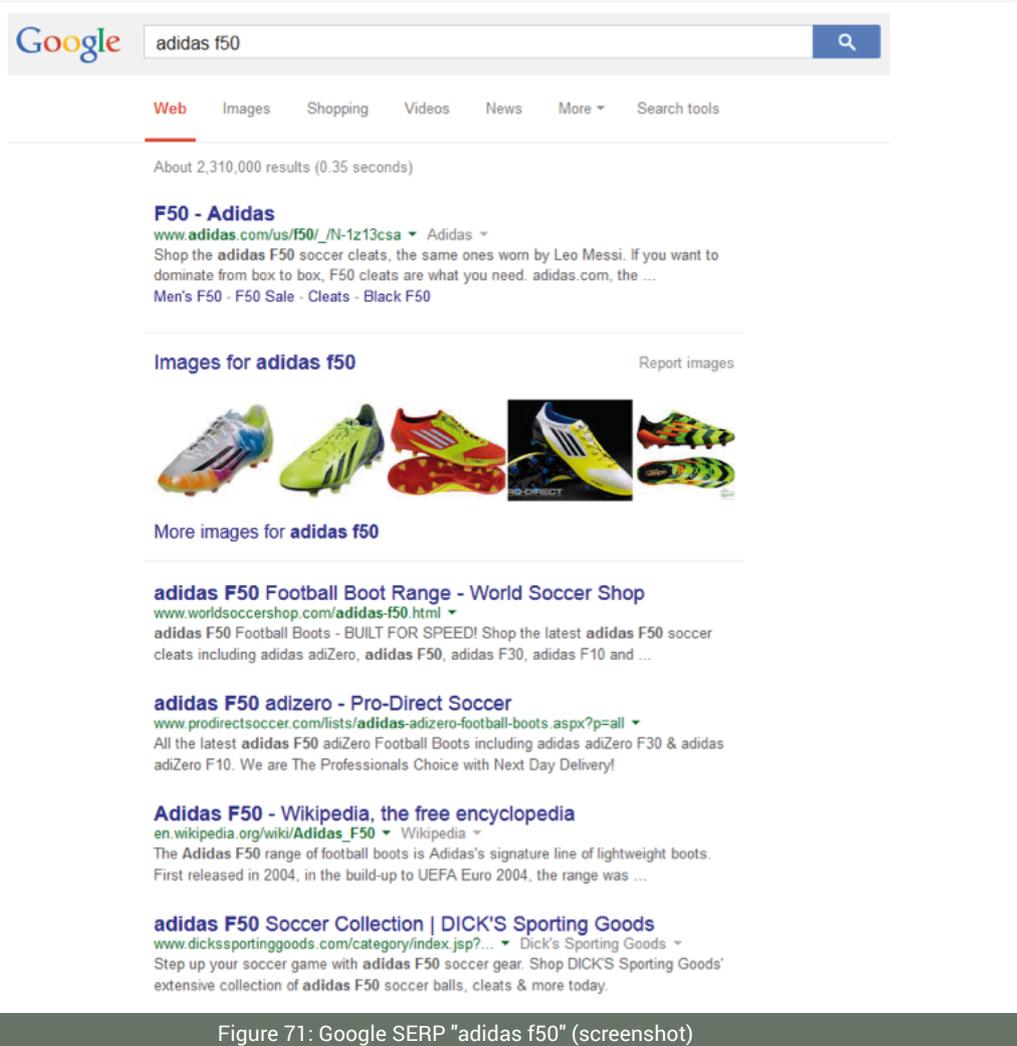
Example: Domains ranking for longtail keywords regarding soccer shoes



2 Big Brands

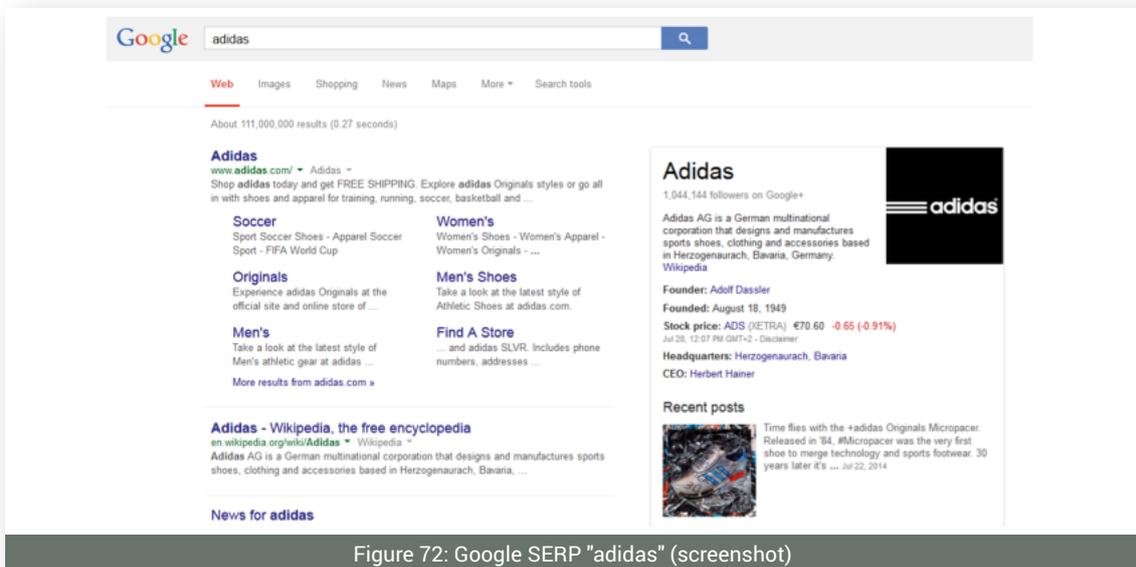
Big brands are characterized by a global presence and a famous brand name. They are often top ranking sites, mainly for generic terms, but also for more specific, long-tail keywords.

Some examples are Adidas.com, Nike.com, IMDB.com, Time.com, WSJ.com, Facebook.com, ebay.com and Amazon.com.



3 Wikipedia ²¹

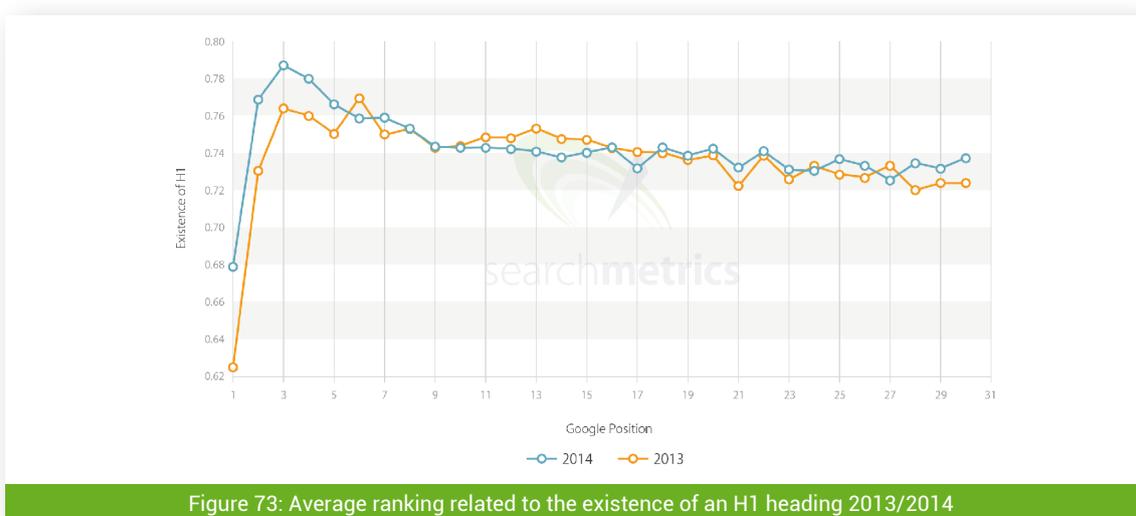
Wikipedia is an exceptional case and can be regarded in principle as a universal Brand. The domain ranks very highly, usually at position 2, for many different short and generic keywords.



Between smaller and larger brands, there are grades of size, with the different gradations being fluid at best. When discussed in this study, the Brand Factor is often a combined basket of all types of brands, as their influence is often difficult to separate out individually.

Brand Factor Has High Impact On SERPs

To further illustrate the impact of the Brand Factor, the graph below shows the anomalies possible with brands in a high search rank position:



²¹ See subitem 'Wikipedia'

The exceptional position for brands can be seen clearly in terms of onpage technical set-up in the above graph. It's clear that brand pages are much less likely to have an H1 heading on the page than any of the other URLs in the top 30 positions in SERPs. The influence of niche brands here is probably significant.

The graph above shows that brands rarely fulfill the „keyword in Description/Title“ factor, but in turn have more backlinks, etc. In contrast to the curve „Existence of H1“ the Brand-Factor includes the „keyword in X“ factor in more positions than just first. It is clear to see the Brand Factor involved here.

This is quite evident from the reports over the past two years. More analysis on the properties of these pages, and on the Brand Factor itself, have been introduced into this year's analysis. An ad hoc starting point was chosen following the observation of obvious anomalies and features.

Search Volume, Pagerank And Alexarank

Additionally this year, other factors that relate in the broadest sense to site quality were introduced, to allow for conclusions to be drawn regarding the Brand Factor, which runs through the entire study.

These features combine content, technology, back link, and traffic features as a classification of web pages based on various factors that also take into account user signals, such as traffic and search behavior.

The selection of these factors was based on the following assumptions:

Hypothesis A: Brands should have a higher search volume. Since it is technically difficult to derive brand names from the pool of existing anchor texts of a domain, the analysis of search volumes has the 'Domain.tld' included in the results.

Hypothesis B: Google's own metric for site evaluation is the PageRank²². Although this metric no longer has its original meaning and/or relevance to the classification of domains/URLs - especially in light of the analysis of backlinks – we should be able to see a positive correlation with rankings and derive an interpretation of the brand factor.

Hypothesis C: Brand Domains should have relatively high traffic - merely by virtue of their good ranking in SERPs. The most common public source of traffic data is Alexa. The Alexa Rank is a figure based on a larger scale that ranks websites according to popularity.

In summary, new features included in this analysis include the search volume of the domain name, including the top-level domain extension, the PageRank of the domain, and the corresponding Alexa Rank. Here the correlation values of these three factors can be seen in detail:

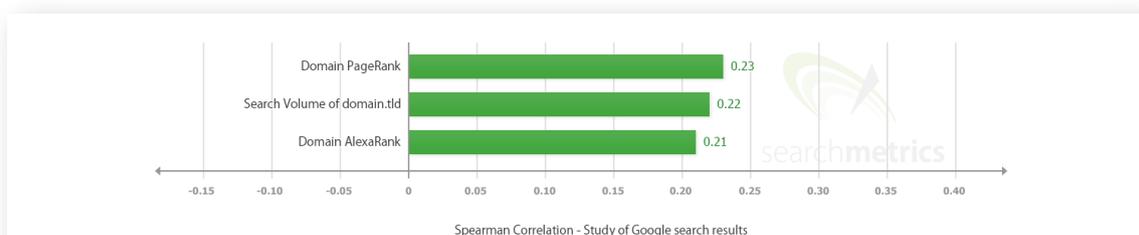


Figure 74: Correlations - Brand factor

²² cf. <http://infolab.stanford.edu/~backrub/google.html>

As you can see, these factors correlate positively with good rankings, and are all in a similar value range. Taking a look at the average curves, however, differences begin to appear.

Note: All features in this range have been calculated without the Wikipedia results.

Pictured below is the PageRank average – an evaluation factor from Google, which, although it no longer has the same importance as it once did, and is not updated as regularly, still remains as a good indication of site quality. Since a high proportion of Deep URLs do not have their own PageRank and the home page is crucial to the Brand factor, the values are calculated based on the respective home page.

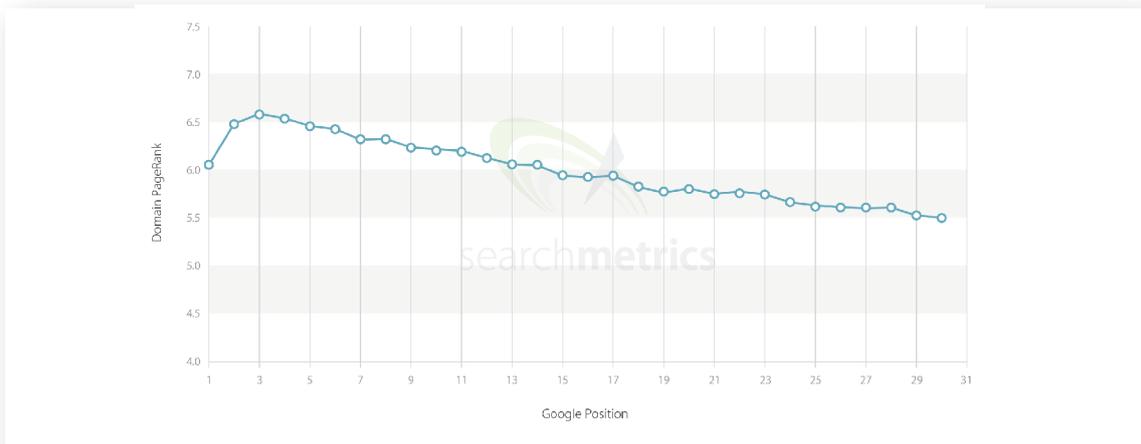


Figure 75: Average ranking - Domain PageRank

The path of the PageRank curve at first appears as you might expect, a clear (positive) correlation in the form of a continuously decreasing curve. The better ranked a URL, the higher their home-page PageRank. Somewhat surprisingly, however, the Brand Factor seems to negatively impact the top positions, with a higher PageRank expected there. This fact can be explained by the presence of Niche Brands, which negatively effect the average here.

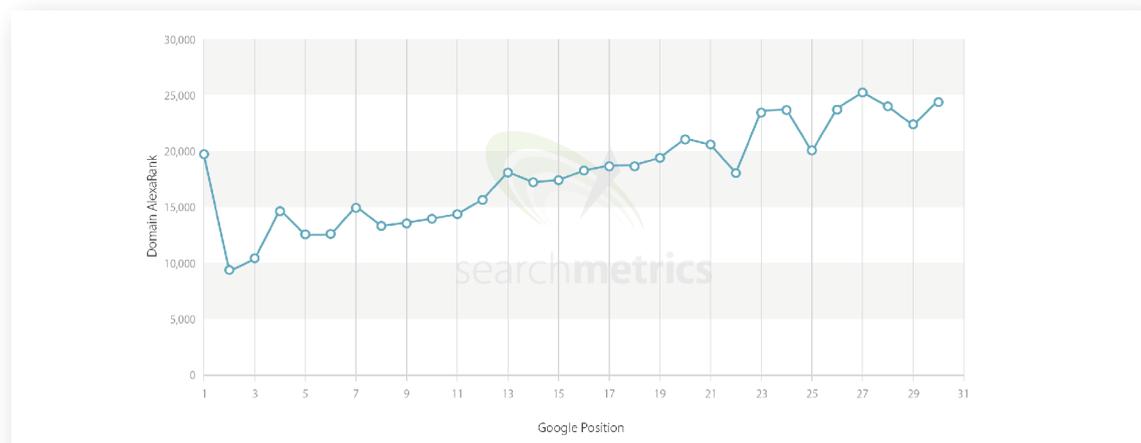


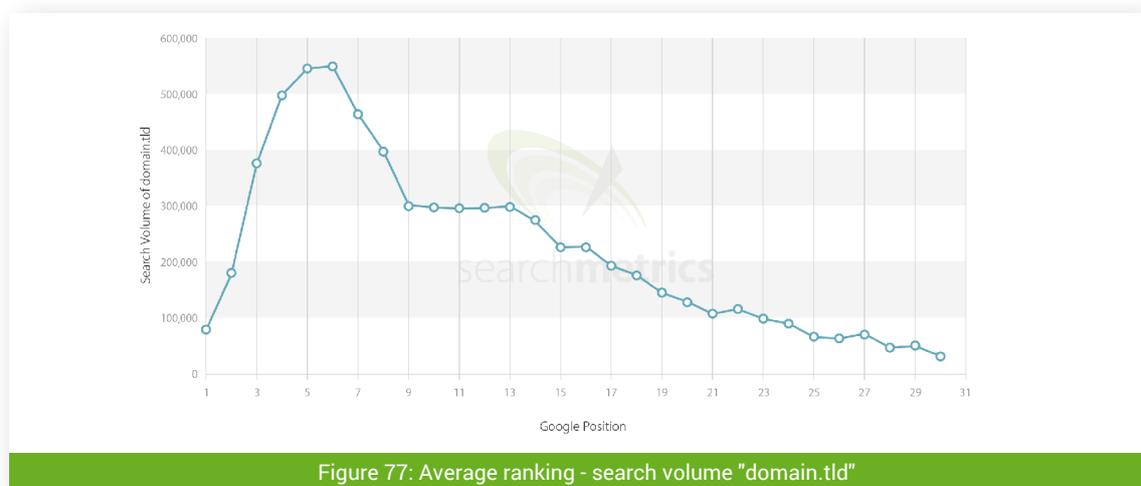
Figure 76: Average ranking – Alexa Rank for Domains

The Alexa Rank graph above shows a mirror image of the PageRank graph. This is to be expected, as the Alexa Rank behaves opposite of PageRank. First, the lower the score, the better, and second, the Alexa Rank is unlimited, whereas PageRank is based on a scale from 1-10.

Alexa Rank is only indicated for the home pages of sites, as opposed to PageRank that covers a wider range of pages²³. This would favor Niche Brands in higher positions, after which the average values fall away slightly before rising steadily.

Homepages of higher ranked URLs have a lower (better) Alexa Rank. The lower a URL is in the search results, the higher the Alexa Rank. Again, this benefits niche brands in ranking positions.

A conclusion regarding the Brand Factor: A URL, or its associated domain, should reflect user behavior or, more accurately, their search behavior. So the search volume of the domain name including the domain suffix (domain.tld), e.g. for „wsj.com“, and its connection with search results has been analyzed this year.



The graphs are similar in the following way: correlations can clearly be seen, and are characterized by a deviation on the top search result positions, as shown here, but the Brand Factor is more pronounced in this case. On average, the „domain.tld“ term in higher-ranked URLs means a higher search volume, except for the top search results, including Niche Brands.

Searchmetrics Page Strength (SPS) For The Homepage

Using Searchmetrics data, we calculated the Page Strength for each domain from a formula taking into account the number and weight of existing backlinks that point to the page. A page with a lot of high-quality links gets a high Searchmetrics Page Strength (SPS) score, a page with few, or no links, gets a small SPS. The calculation is strongly geared toward the PageRank formula.

The correlation for SPS was calculated based on the respective home page of each website.

²³ Therefore, the values of the homepage of a URL were evaluated for both Pagerank and for Alexarank

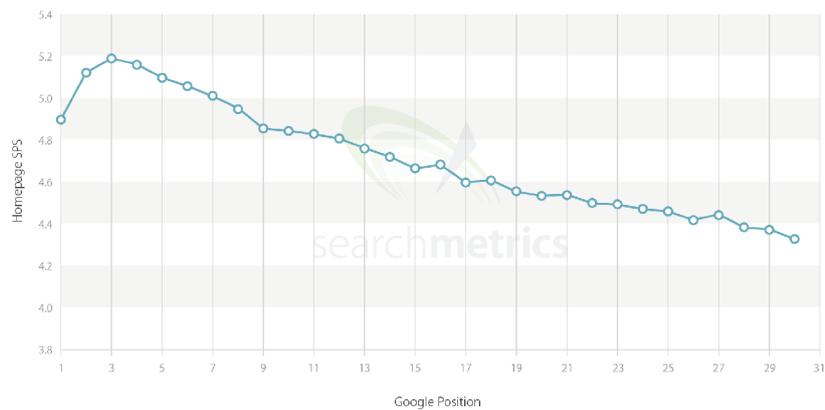


Figure 78: Average ranking according to - "Searchmetrics Page Strength" for homepages.

The curve is very similar to that of the correlation for PageRank, which was expected due to the basis of the calculation. It can therefore be said that Searchmetrics data supports the correlation with the PageRank rankings, both in fact and in interpretation.

„Brand Links“ - What Is A Brand And Are Anchor Texts A Brand-Reference?

Based on the method of analysis of Brand-terms for domains based on the frequency of anchor texts, a correlation for the proportion of Brand links and search result positions was calculated:



Figure 79: Average ranking - share of Brand links

This graph shows the proportion of Brands on position X - or more precisely, the percentage of URLs where the domain name (without „tld.“) is the most used anchor text on all links. The analysis resulted in a positive correlation with improved rankings.

In addition to the influence of niche brands on the average values in the top positions, further explanations of the fall-off at position 1 are required. For Brand sites, the proportion of Brand links may be lower because they are linked with different anchor texts (also generic) or images, and they

generally have more links. However, the absolute number of Brand links is likely to be higher and the proportion of all links lower.

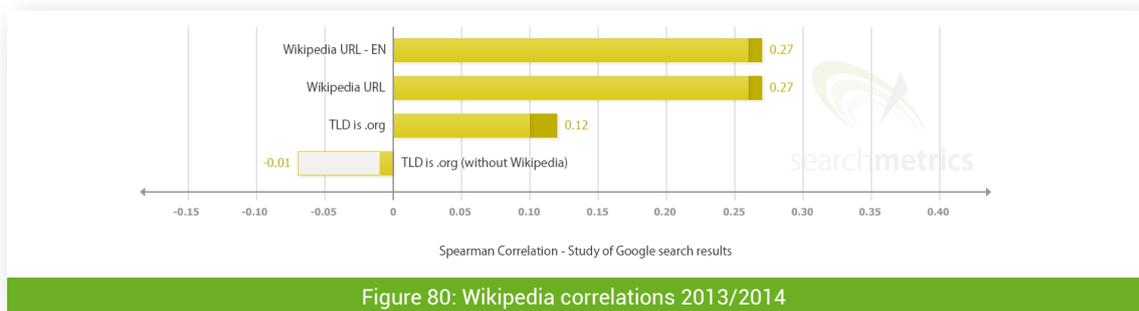
In addition, there are usually several Brand names for domains, some of which differ only typographically. In the graph for anchor texts for the domain people.com there are four Brand-Anchors („people“, „people magazine“, „People“ and „People Magazine“) whose metrics could potentially be added. Technically, however, only the extraction of Brand-Anchors, according to the pattern above, was possible for this analysis. Consequently, the curve should probably be different for the top positions and the correlation would then be positive.

Wikipedia: The Best-Optimized Universal Brand

Wikipedia has a significant influence both on the ranking distribution in search results, and on the correlations calculated for this study, but not in the way that it had in 2013.

Some of the Wikipedia rankings in the previous analysis influenced the results to such an extent that, for a few calculations, it was necessary to filter the results in order to make accurate predictions for the rest of the search results. Graphs for individual factors identified variances at positions 17 and 27. These were exactly the positions where the Wikipedia results had ranked for a long time.

This was not observed in 2014, but the correlation for „Wikipedia features“ has risen again.



The graph shows that a relatively trivial statement such as “URL from Wikipedia” is definitely an important ranking factor.

In addition, on the basis of two studies on the domain extension “.Org,” we show how strongly the correlation for „TLD is .Org“ is influenced by Wikipedia.

The crowd-sourced project benefits from an extremely high degree of trust and ranks for many keywords (especially „Informational Keywords“) at the top positions in SERPs.

It could be stated that the typical Wikipedia entry²⁴ is on position 2 behind the Brand site or the URL for the specific term search being conducted. It is often the case that Wikipedia ranks on position 1 where no significant brand is related to the keyword search term, especially if it’s an „Informational Keyword“.

Just how strong the impact of Wikipedia is on search results can be observed from the following detailed correlation of the „URL is from Wikipedia-EN“ feature.

²⁴ Taking into account the search volume of a keyword and thus the query frequency

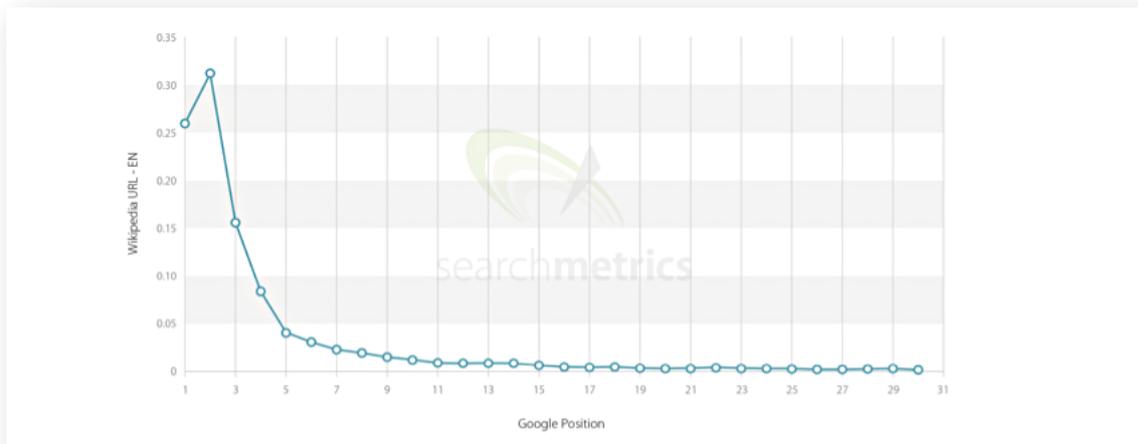


Figure 81: Average ranking - Wikipedia URL

The analysis shows that Wikipedia URLs rank very highly on Google.com, mostly on positions 1 or 2, but sometimes at 3.

The clusters on positions 17 and 27 that were apparent in the 2013 analysis no longer exist, due to the Penguin 2.0 update in calendar week 22/2013.

Wikipedia content does not rank in the same way that Google wants it to. In principle this is the most optimized website in the world. It has continually updated content, its relevance is continuously checked and content-clustering and holistic content can be found naturally on almost every page. The domain is rigidly and relevantly linked internally and has a natural, ever-growing link profile. The whole concept is to establish an international presence - in other words to be in every respect a Universal Brand.

The Top Positions Are Reserved For Brands - And Wikipedia

Brands rank in the top positions even if they don't meet certain criteria at all, or at best, meet them loosely. Here, a distinction should be made between smaller, niche brands, larger brands that are known globally, and Wikipedia – THE Internet Brand. It is our conclusion that Google treats Brands in a unique way, and they are often given the best possible reward, the top position in SERPs.

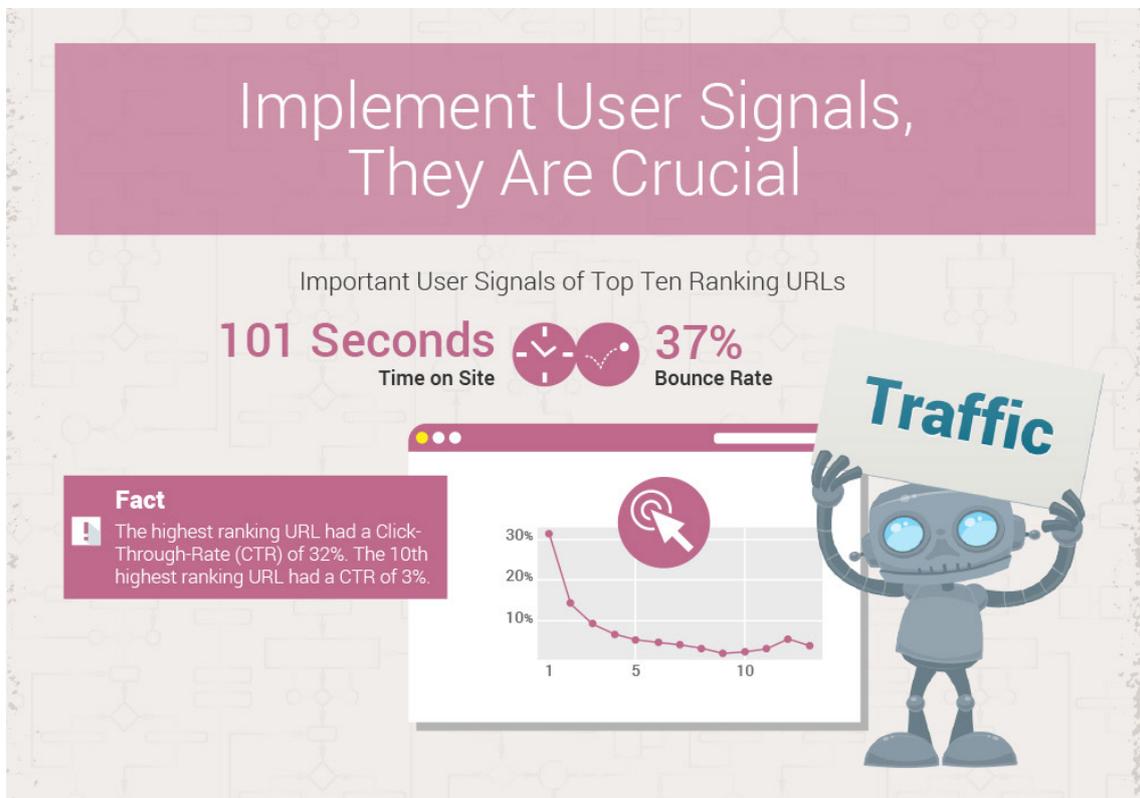
For non-Brands, the requirements needed to rank on position 1 are very high, with the top search position for certain search queries seemingly reserved for brands even when certain ranking factors are not met completely, or are totally absent.

These missing factors, which are a pre-requisite for non-brands to stand a chance of ranking highly, are mainly in the onpage range. Brand pages, for example, have the keyword in the Title or Description more rarely and on average have less content and fewer internal links.

However, what seems to impact for brand pages very positively is an exceedingly good backlink profile. Brands are also frequently linked to sources with high SEO visibility. In conclusion Brands receive appreciably more social signals and offpage factors significantly influence the position of brands.

As a universal Brand, Wikipedia is given ranking priority by Google, and deservedly so. Wikipedia has constantly updated, holistic content, a well structured site architecture, and is relevant for many keywords in the search results.

At a glance: Infographic – User Signals



6. USER AND TRAFFIC SIGNALS

The correlations detailed for the preceding factors are all based on the analysis of freely available data - actual, existing rankings. However, we are dealing with a bird's eye view in this case. Only Google knows the actual algorithm, and thus the true ranking factors.

To enrich our findings, we investigated user traffic signals using our own internal and unique data. This data was first made anonymous and averaged. The data used was based on a different keyword set than used for previous calculations, due to the amount of data available, and is valid for the top 15 positions in the SERPs.

First, a look at the correlations:

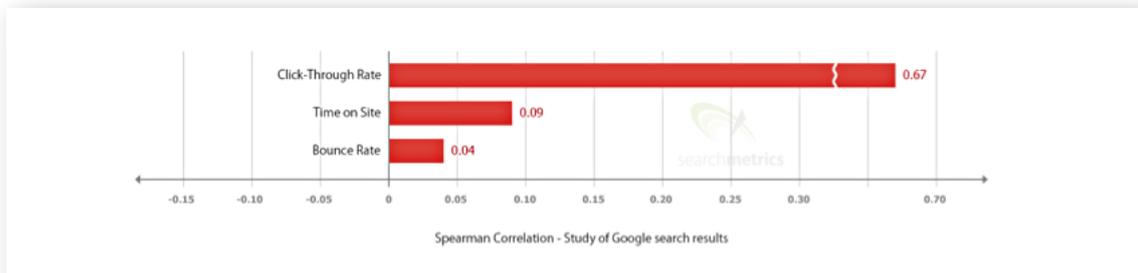


Figure 82: Correlations - User-signals 2014

If the correlation for „click-through rate“ was part of the overall bar chart, it would be right at the top. This is the highest correlation that has ever been calculated for a factor of ranking studies.

The basis for this analysis is the average CTR and we had to process the data differently than we do in the Searchmetrics Suite. For the calculation of SEO Visibility and other metrics in the Searchmetrics Suite a dynamic CTR curve is used, which is quite complex. For the present study, the average of all CTRs was analyzed, even if the CTR per keyword and search intention were quite different. In detail, the curve looks like this:

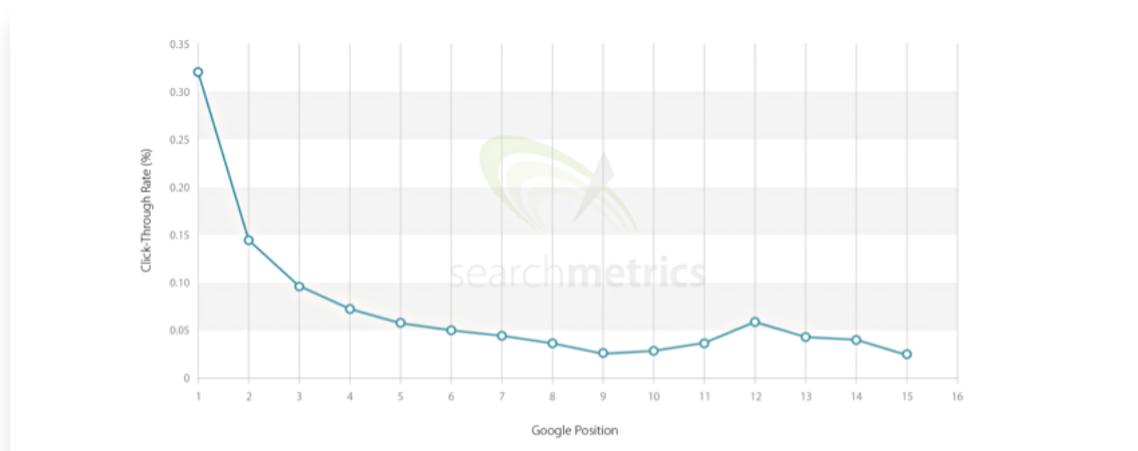


Figure 83: Average ranking - CTR

It is possible to see an exponential curve, with the top search results clicked more often. This is not surprising, and makes absolute sense.

If the user clicks, they stay longer on pages that are higher ranked. This is shown in the time-on-site metric:

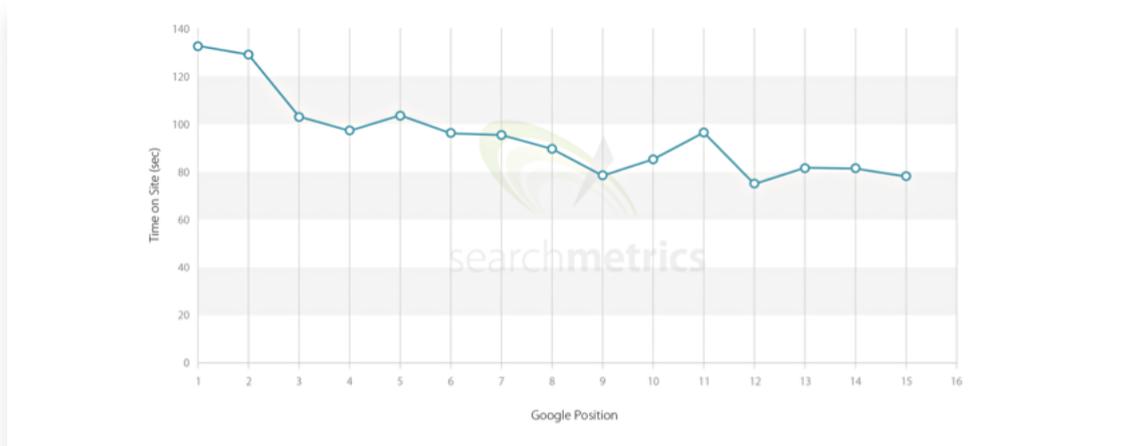


Figure 84: Average ranking by Time-on-Site

Users who click on the first result stay on the page, on average, 30 seconds longer than those who click on a page on position 4. This trend continues as one moves down the rankings.

But it should be noted that the visit time also depends on the type of search and the result. The motivation and purpose of the user for certain information will shape the visit, for example when looking at a sports result from the night before, the visit will usually be very short. Here again we have formed averages over all keywords.

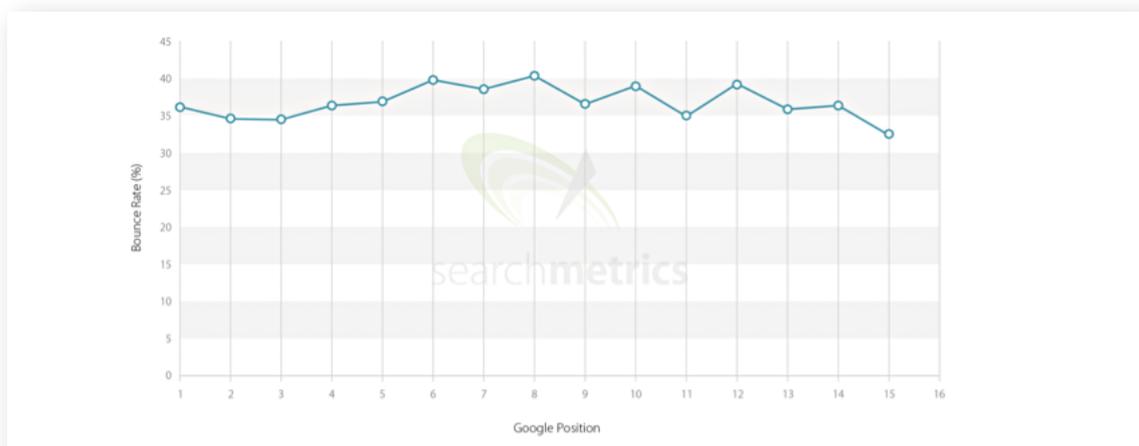


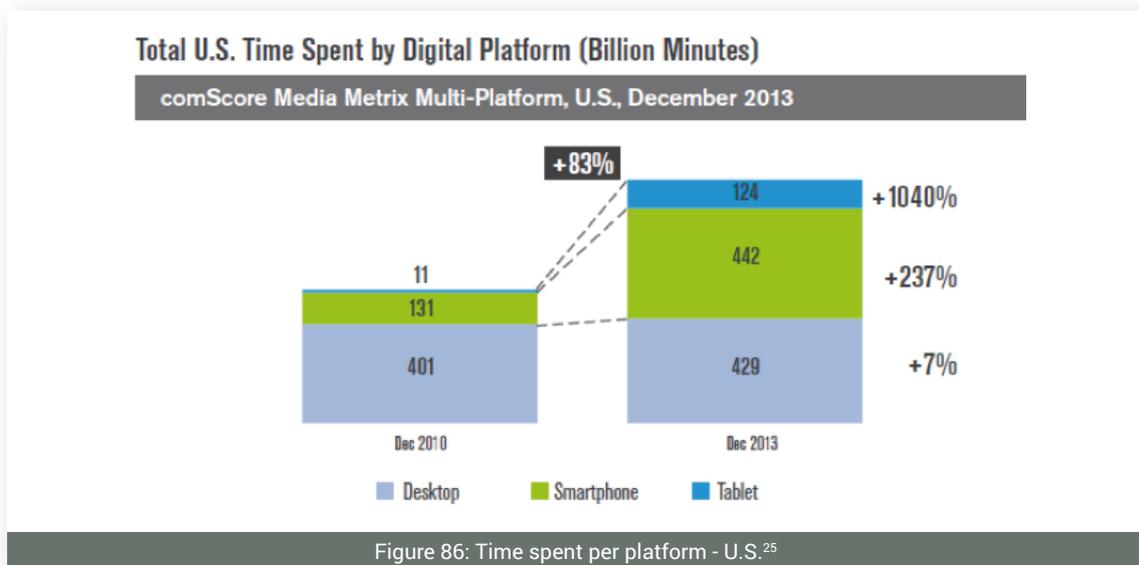
Figure 85: Average ranking according to bounce rate

Just as for time-on-site, the bounce rate is a strong quality signal. Users that bounce – click on a site and then leave it quickly – are assessed to have not found the right information for the result of their search query.

The Bounce Rate is another metric that can be interpreted according to behavior. A hypochondriac, for example, may call on 'Dr Google' for advice relating to self-diagnosed symptoms by searching for medical information online. Even if the user is pleased with the content of one website, they very often want further information to develop an overall understanding, or to obtain several opinions. Their visit pattern would be one of bounce after bounce, but this behavior shouldn't be interpreted negatively.

7. DIGRESSION AND OUTLOOK: DESKTOP VS MOBILE

Search is changing, and changing rapidly. Google is constantly testing their own performance and revising the search results and their algorithm. But that is only one side of the coin. User behavior is also evolving.



„... The most significant development in the search marketplace right now is the continuing shift to mobile platforms... as people now spend more time accessing the Internet on their mobile devices rather than desktop computers. An increasing number of searches are conducted on smartphones and tablets²⁶.“

The time spent by users accessing the web via mobile platforms is growing every day, and with it, the number of mobile searches carried out. This is not just a US phenomenon.

In addition to the trends toward Contextual and Semantic Search, the Mobile Search area (including Conversational Search) is currently one of the most important features of user integration on the web. Mobile searches often use different functionality and are strongly influenced by features such as conversational (voice-based) search with Siri, Google Glass, etc., as well as being influenced by local factors.

There are differences between mobile and desktop search results, and the extent of these differences was examined by Searchmetrics for a representative keyword set for Google.com (U.S.). The findings are shown below:

²⁶ comscore.com/Insights/Presentations-and-Whitepapers/2014/2014-US-Digital-Future-in-Focus, p. 34

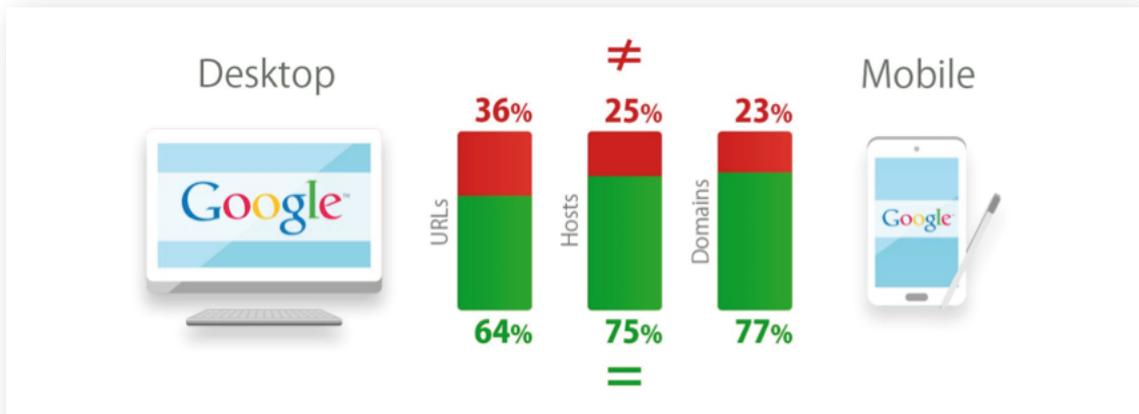


Figure 87: Desktop vs. Mobile - Differences for URLs, hosts, domains in percent (google.com)

Not only are the SERPs for mobile different, but the URL page rank properties differ too.

In a study on specific mobile ranking factors for this analysis, Searchmetrics has investigated the differences between these two areas. Here are two examples:

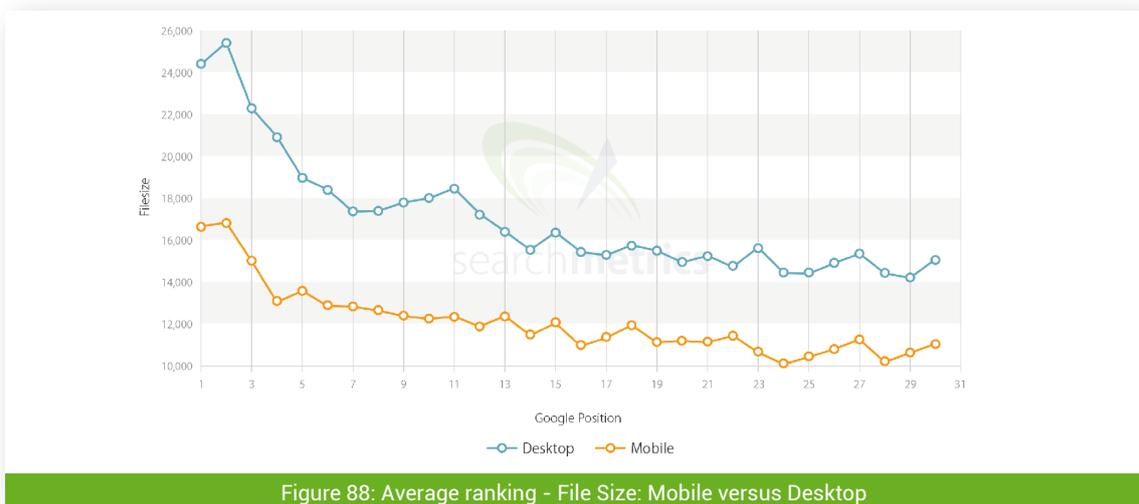


Figure 88: Average ranking - File Size: Mobile versus Desktop

Mobile search results for the top 30 positions return, on average, a much smaller file size than for desktop. This is logical, as content for mobile-optimized pages is often shorter. From a user's perspective, it makes sense to view smaller data packets, not to exceed certain data volumes and to avoid longer download times.

Mobile optimized sites still climb the rankings, even with a very different backlink profile:

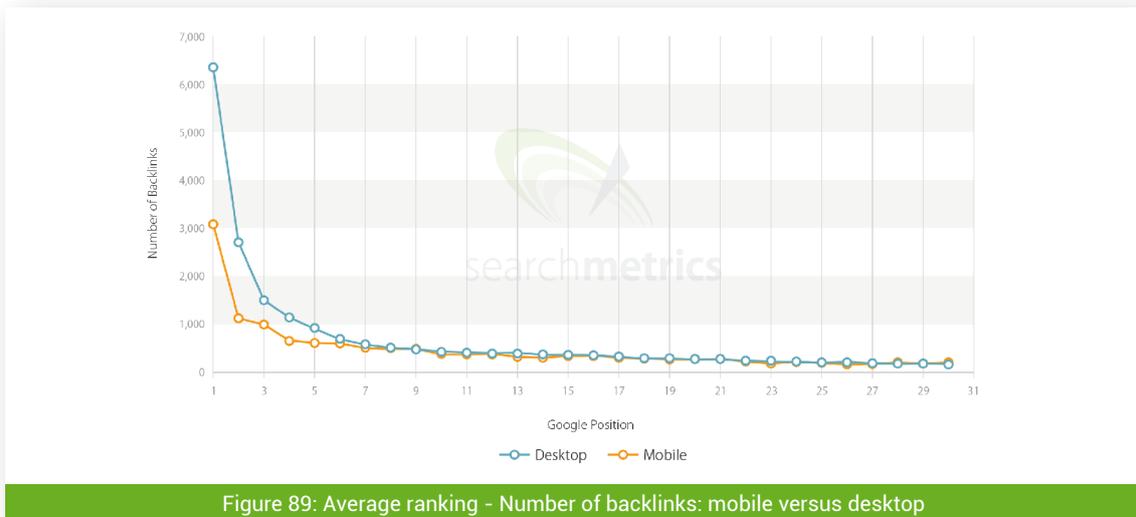


Figure 89: Average ranking - Number of backlinks: mobile versus desktop

Mobile search results for the top 30 positions have, on average, fewer back links compared to desktop rankings. Given mobile user behavior this is understandable. Mobile and tablet content often carries fewer ad hoc links, and this is one possible explanation for the differences shown above.

Social signals for mobile look very different, too, and more on this is to follow in the Searchmetrics study on mobile ranking factors and rank correlation, to be released later this year.

Summary of Searchmetrics SEO Ranking Factors 2014

To conclude this study, we present an overview of the totality of all rank correlation coefficients and the respective changes compared to 2013. Factors new for 2014 are presented as grayed-out in Figure 93.

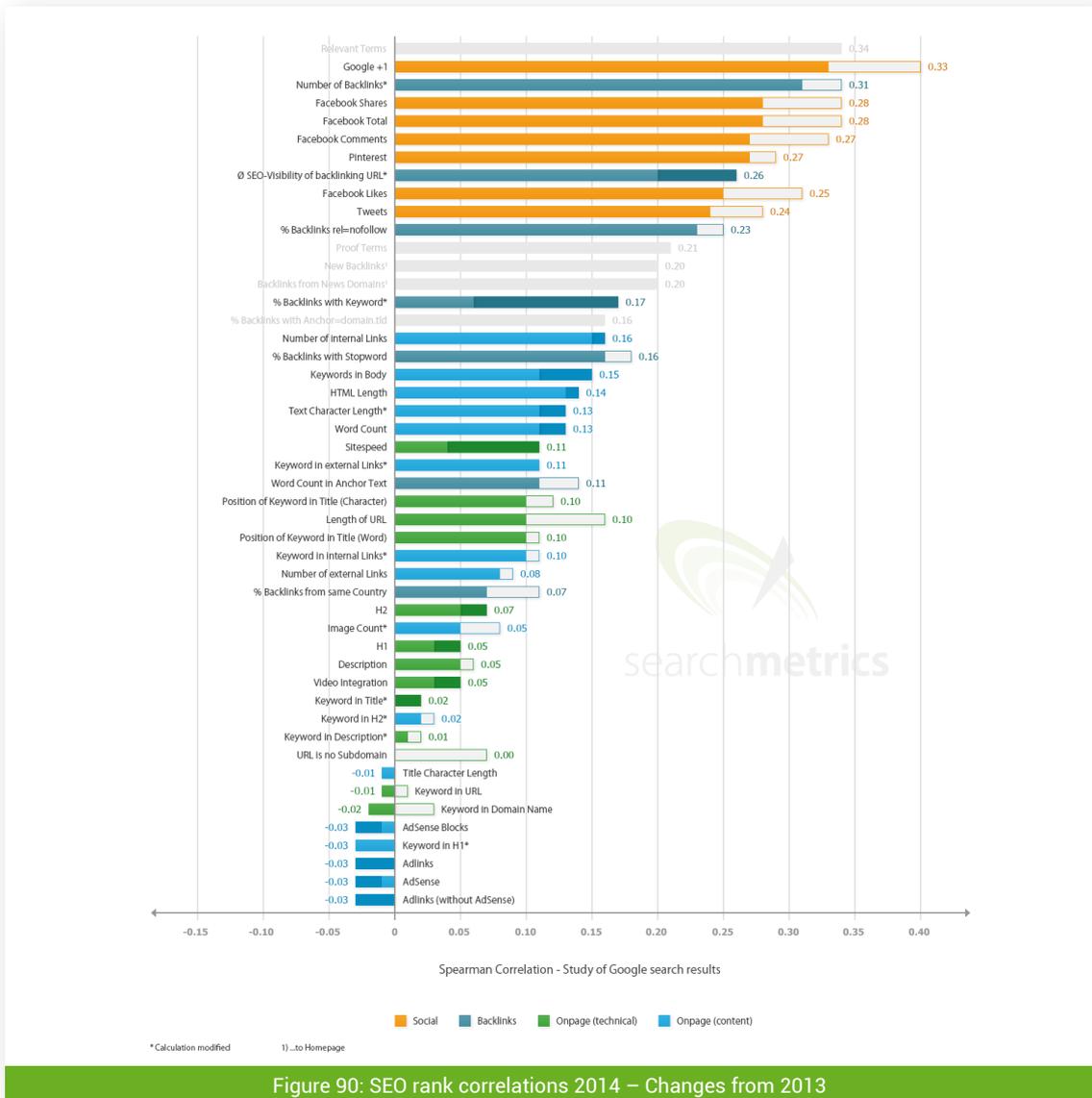


Figure 90: SEO rank correlations 2014 – Changes from 2013

Summary: The Main Developments For Ranking Factors In 2014

Content is no longer an addition to, but is the main focus of, SEO. For a long time this was a difficult goal to achieve technically (from a search engine perspective) yet it was still easy to be manipulated (from an SEO perspective). Recent developments, both to the Google algorithm and with the impact of each specific update, clearly shows that relevant, holistic content ranks better over the long-term.

The content basics are certainly keyword term frequency and co-occurrence analyses. Consequently, keywords also need to be expanded upon in the copy, with reference to other semantically relevant terms. Focusing on a keyword together with other semantic terms can be effectively optimized, depending on the content and ranking intention. In fact, better ranked URLs have, on average, more holistic content.

There has been a general rise in content word count, HTML length, and related features. Furthermore, additional media, such as images, both enrich the content, and are likely to increase the relevance of a search result, especially when user signals are added in for measuring the interaction with these types of media. For English language pages, better ranking content is easier to achieve when it is easier to read. Pages in the top 30 positions have, on average, less advertising on the site - and Adlinks of any kind in top ranking positions are rare.

Good housekeeping is very important; strong information architecture is key, and poor internal link structures can have a negative impact on search engine performance.

Up-to-date site and page technology is a basic requirement for a good ranking, including loading times and the presence of all common Meta elements and H (x) tags. Keyword domains are found even less frequently in the SERPs than in the previous year. The presence of a keyword in the URL remains in some way „natural“, but is not a ranking factor.

For onpage keywords, things have changed. On the one hand, it should be clear by now that keyword stuffing is history, but on the other hand it's natural that keywords and related search terms should occur frequently in the content copy.

Social Media is, to put it mildly, a small „loser“ in this study, with slightly decreasing correlations and only a small growth in averages for the number of Likes, +1s, tweets etc. by position.

Backlinks, by contrast, remain important, with emphasis on their quality, much more than their quantity. Anchor text keywords on top-ranked pages are significantly less common than in 2013. In terms of the proportion of nofollow links, stop words in the anchor text, and the number of words in the link text, Google seems to have found a good blend. These numbers have not changed much compared to the previous year.

In order to interpret this year's new features and findings, a new definition of Brand was needed. Generally, bigger brands rank among the top results, but not necessarily at position one, where smaller brands regularly take the top spot. Bigger Brands have more news links in general, more links from more different domains, more links from news domains, etc. For all these factors there were positive correlations based upon average values, and these differed only at position 1 (often smaller brands) and 2 (often Wikipedia).

Lastly, healthy user signals are increasingly important for good rankings, and our analysis shows Google seeming to favor their impact on search results. Top-ranked searches have a significantly higher CTR, as well as a lower bounce rate, and a longer time-on-site. These are metrics that interact strongly with each other and seem to greatly influence SERPs. After all, it is essential to remember that search results try to reflect one thing, user intention.

Over all, the principle remains that good search ranking positions generally cannot be achieved by cherry-picking a few factors. Having many backlinks and a fast load-speed will not result in a high ranking if the content on the page is not relevant to the user.

Good rankings are based on the interaction of many different and differently weighted factors.

For questions regarding this White Paper, the Searchmetrics Suite™
or to find out more about how we can help your business, contact:

sales@searchmetrics.com

For more information, please see our website:

www.searchmetrics.com



Searchmetrics

1510 Fashion Island Blvd, Suite 250

San Mateo, CA 94404

Phone: 1 866 411 9494

E-Mail: sales-us@searchmetrics.com

