



WHITE PAPER

SEO Ranking Factors – Rank Correlation 2013

- Google USA -

Executive summary:

This white paper deals with the definition and evaluation of factors that have high rank correlation coefficients with top organic search results and aims to provide a more in-depth analysis of search engine algorithms. The data collection and its evaluation refer to Ranking Factors and Rank Correlation for Google.com in 2013.

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

About Searchmetrics and the Searchmetrics Suite™

Searchmetrics is a pioneer and the world's leading provider of search and social analytics software for companies and online agencies.

Our software, Searchmetrics Suite™, empowers users to increase visibility, market share, revenue and profits by applying optimal strategies in the digital marketing world.

The main advantages at a glance:

- ✓ **Identification** of competitors and their strategies
- ✓ **Optimization** of the online marketing mix
- ✓ **Increase** in the company's ROI
- ✓ **Largest international and local coverage** in the world, with 124 countries and 155 local rankings, allowing for the optimization of global or region-specific search campaigns
- ✓ Monitoring of brand names in **268 search engine/country combinations worldwide**
- ✓ ROI analyses: **Discovering untapped potential** by displaying keywords with the corresponding search volumes and CPCs

Would you like to find out more about Searchmetrics Suite™ or request a quote?

Contact us now!

Contents

FINDINGS OVERVIEW	5
Structure of the white paper	7
DEFINITIONS, INFLUENCING FACTORS, AND THE DATA POOL.....	8
What is a Ranking Factor?	8
Ranking Factors: Correlation \neq causation	9
Influencing factors.....	10
Data freshness.....	10
Data pool	10
Interesting background facts	11
The correlation coefficients at a glance.....	12
1. ON-PAGE CODING	14
1.1 On-page coding: Developments since last year.....	15
1.2 Keyword domains are no longer a Ranking Factor	16
1.3 Presence of keywords in the description, title and H1 remains important ...	20
1.4 WDF*P*IDF	21
1.5 Keywords in the title should be placed as close to the front as possible	23
1.6 Page load times as an important performance factor	23
1.7 DIGRESSION: Authorship integration does not (yet) have an effect	24
1.8 Summary.....	26
2. ON-PAGE CONTENT	28
2.1 On-page content: Developments since last year	29
2.2 Higher number of words in the text is – in contrast to the previous year – clearly important.....	31
2.3 HTML length is important but only to a certain extent.....	32
2.4 Text length is relevant only up to a certain point	33
2.5 Using media to enrich content pays off	34
2.6 Internal links: Number, structure and the text of links optimize link juice ..	35
2.7 Moderate integration of AdSense and other Adlinks is less negative than before.....	38
2.8 Summary.....	41

3. BACKLINKS.....	42
3.1 Backlinks: Developments since last year.....	43
3.2 The number of backlinks remains immensely important.....	44
3.3 Link diversity is always relevant	44
3.4 "Hard" keyword anchor text has lost its relevance.....	46
3.5 Backlinks from URLs with high SEO visibility are significantly more valuable.....	48
3.6 Combinations with backlink features	50
3.7 Summary.....	50
4. SOCIAL SIGNALS	52
4.1 Social Signals: developments since last year	53
4.2 Social signals continue to show strong correlation with good rankings	53
4.3 Google+ new in the analysis - with very high correlation	56
4.4 Digression: Google+ will overtake Facebook at the beginning of 2016	57
4.5 Summary.....	59
5. THE BRAND FACTOR.....	61
5.1 High impact of the brand factor on SERPs	61
5.2 The top ranks are reserved for brands.....	62
6. THE SPECIAL ROLE OF WIKIPEDIA.....	64
6.1 Wikipedia results affect correlations	64
6.2 Rank 17 and 27 "reserved" for Wikipedia results.....	66
SUMMARY OF SEARCHMETRICS SEO RANKING FACTORS 2013	67
Summary of most important developments in ranking factors 2013:.....	68
ABOUT SEARCHMETRICS	69

List of Figures

Figure 1: Searchmetrics – SEO Ranking Factors – Rank Correlation 2013	12
Figure 2: On-page coding – Rank Correlation 2013	14
Figure 3: On-page coding 2013 – Changes since 2012	15
Figure 4: Correlation – Keyword in Domain / URL	16
Figure 5: Screenshot – Losers at keyword level for an example domain after Penguin update	17
Figure 6: Correlations – Existence of on-page keywords / description, H1 and H2	20
Figure 7: Average by ranking – Similarity	22
Figure 8: Correlations – Position of the keyword in the title	23
Figure 9: Average by ranking – Page load time	24
Figure 10: Average by ranking – Presence of authorship integration	26
Figure 11: On-page content – Rank Correlation 2013	29
Figure 12: On-page content 2013 – Changes since 2012	30
Figure 13: Correlations – Content quality	31
Figure 14: Average by ranking – Word count in text	32
Figure 15: Average by ranking – HTML code length in characters	33
Figure 16: Average by ranking – Text length in characters	34
Figure 17: Average by ranking – Image count	35
Figure 18: Average by ranking – Number of internal links	36
Figure 19: Average by ranking – Number of internal links pointing to the same website	37
Figure 20: Correlations – Ad integration	38
Figure 21: Average by ranking – AdSense	39
Figure 22: Average by ranking – Adlinks total (including AdSense)	39
Figure 23: Average by ranking – Number of AdSense ads	40
Figure 24: Backlinks – Rank Correlation 2013	42
Figure 25: Backlinks 2013 – Changes since 2012	43
Figure 26: Average by ranking – Number of backlinks	44
Figure 27: Average by ranking – number of words in anchor text	45
Figure 28: Average by ranking – Percentage of nofollow backlinks	46
Figure 29: Average by ranking – percentage of backlinks with keyword	47
Figure 30: Average by ranking – Percentage of backlinks with stop word	48
Figure 31: Average by ranking – Average SEO visibility of the link source	49
Figure 32: Social signals – Rank Correlation 2013	52
Figure 33: Social signals 2013 – Changes since 2012	53
Figure 34: Average by ranking – Facebook signals in total	54
Figure 35: Average by ranking – Twitter signals	55
Figure 36: Average by ranking – Pinterest signals	56
Figure 37: Average by ranking – Google+ signals	57
Figure 38: Percentage growth and absolute number of Facebook shares vs plus ones	58

Figure 39: Facebook shares vs. plus ones – growth forecasts in %	59
Figure 40: Average by ranking – H1 title tag.....	61
Figure 41: Average by ranking – Keyword in the description	62
Figure 42: Wikipedia correlation.....	64
Figure 43: Average ranking – Wikipedia URL.....	65
Figure 44: SEO Ranking Factors – Rank Correlation 2013 – Changes since 2012	67

Findings Overview

On-page factors such as the technical side structure and good content are basic requirements for good rankings. The keyword on-page is important, but seems to have been devalued by Google in its relevance as a domain name, as a part of the URL and as the link text.

In contrast, the number and quality of backlinks are probably very important – "natural link profile" and diversity are the keywords here. Social signals continue to correlate positively with better rankings.

Structure of the white paper

The study is focused, on one hand, steps involved in building a website in practice and on the other on the relevance of influencing factors, which has increased over time.

The study first examines the on-page area – broken down into two thematically different blocks, *coding* and *content*. This is followed by the analysis of the off-page area in the form of *backlinks* and *social signals*. The white paper concludes with the analysis of two factors that play a special role in this study and therefore have to be examined separately. The final summary of the findings of the study follows the relevant chapters on the brand factor and the role of *Wikipedia*.

The chapters first present an overview of all factors influencing the area concerned followed by the trends in the development of factors in 2013 examined in relation to the previous year. At the core of every chapter are the remarks and conclusions on the individual factors, followed by a chapter summary at the end of each chapter.

In addition, the impact of correlations on a combination of factors is discussed in individual chapters.

Definitions, influencing factors, and the data pool

What is a Ranking Factor?

Search engines use algorithms to rank websites by subject and relevance. All the websites in a search engine index are categorized based on this structure, which finally results in an optimal ranking for user's search queries. In general, the criteria for evaluating websites and performing this ranking are referred to as Ranking Factors.

The reasons for this are obvious: Due to the exponential rise in the number of documents on the Internet – and thus in the search index – it would not be possible to rank websites without automatic algorithms despite the existence of human quality raters. While, on the one hand, these algorithms are trivial (order requires a pattern), they are also the best-kept secret in the Internet business, because it is essential for search engine operators to keep the underlying factors influencing the algorithm strictly confidential.

The reason for this is elementary rather than competition-related: If the definition and the degree of influence of factors relevant for a good ranking were known, it would render them at the same time irrelevant – because then they could be manipulated.

At the beginning of the search engine era, Google would rank websites on certain topics as relevant based on the high keyword density on these topics. However, webmasters used this knowledge to climb up the SERP ranking by using keyword overloaded pages that resulted in irrelevant searches¹.

This fact has not only triggered real competition between search engines and website operators, but also the mystery surrounding the Ranking Factors. Aimed at developing semantic search, a set of criteria was devised, which was initially strictly technical (e.g. number of backlinks), but in the meantime also contains less technical components (e.g. user signals).

This fact, combined with the pursuit of optimal search results, culminated in the continual evolution of Ranking Factors, whose categorization and complexity due to ongoing iterative update cycles under the strong influence of user signals is subject to constant optimizations aimed at producing the most relevant search results for users.

For website operators this means a steadily decreasing number of factors that can be negatively manipulated, on the one hand, and the opportunity – given the increasing efforts to combat spam, irrelevance and short-termism – to achieve

¹ **Search Engine Results Pages**

high search result rankings in the long term using a sustainable business strategy based on relevant quality factors, on the other.

Ranking Factors: Correlation \neq causation

We are not Google. The analysis and evaluation of Ranking Factors using this data not only has interpretation value but, in fact, represents a profound interpretation (and thus not a mere conjecture) on the basis of facts, namely the evaluation and categorization of website features that achieve high ranking in search results.

We at Searchmetrics aggregate billions of data points every month and naturally look for the answers to the question: Which factors are relevant for a good ranking in Google's search results?

To this end, we compare website characteristics with their Google rankings and derive a structured list of Ranking Factors. For example, if there are many pages in the top positions of analyzed SERPs that contain a keyword in the title tag, whereas fewer low ranking pages have that feature, then we have identified a high correlation with good rankings. On the basis of these correlations we can then draw conclusions about the characteristics of websites that rank well.

However, it is necessary to point out that correlation is not synonymous with causation, and thus there is no guarantee that the relevant factors actually affect rankings or are even used by Google as signals. For this reason, Searchmetrics conducts studies on a regular basis which analyze, for example, the impact of social signals under isolated conditions. As a result of this, the impact of social signals, among other factors, on rankings has already been proven.

Nonetheless, the data speaks for itself, unlike the definitive Ranking Factors and their degree of influence, this data is not a secret. As a result, our analyses allow us not only to assess what factors will most likely be included in the ranking, but also the degree of their influence. Many observations, which were previously not based on much more than a gut feeling, can now be confirmed with the help of the following data.

As already mentioned, there are purely technical or "hard," as well as less technical or "soft" Ranking Factors. In order to optimize the visual impact of our charts, we have grouped our factors – even if some areas partly overlap - into *social, backlinks, on-page coding and on-page content*.

As the term "Ranking Factor" and its causal elements of meaning can lead to problematic interpretations, we prefer this term to be rather interpreted as rank correlation coefficient in the context of our analysis and this white paper.

Influencing factors

The position and order of search results depend not only on the entered search queries (= keywords), but also personal search history, which in turn is determined by the behavior of the user. In addition, it depends on factors such as: Is the user logged into Google? Does the user use Google's own Chrome browser to surf the Internet? Where is the IP of the user located? Are cookies containing personal information stored on the user's computer, etc.? Last but not least, Google itself is changing the variations of these integration types all the time.

Since the consideration of these many individual factors would make the analysis excessively complex, this white paper focuses on the variations of the integration types introduced by Google itself on a basis that is as neutral and national as possible.

Data freshness

The data was first collected and aggregated in March 2013 and then again, in June 2013². As this is a follow-up study to the 2012 Ranking Factor analysis, the changes of many factors compared with the previous year are particularly interesting. They provide invaluable insights into the evolution of Ranking Factors, which can be used to derive trends for future development and the relevance of the criteria. The differences compared with the previous year are considerable in some areas.

In addition, the pool of analyzed characteristics was expanded to include many factors that allow more in-depth analyses of existing factors or shed light on new fields.

Data pool

These analyses are based on search results for a very large keyword set of 10,000 search terms from Google USA. We have largely excluded specific navigational keywords (as it was the case last year) from the original data pool, which included the top 10,000 keywords according to search volume, as we did not want to distort the analysis. Search queries are considered to be navigational keywords if they return results, which are largely irrelevant and only return one exact match (for example: "Facebook Login").

The first three pages of organic search results were always used as data pool. This keyword set covers approximately 90 to 95 percent of the data pool used in the study last year. Here, we attempted to strike a balance between two factors,

²Second data point: after the Google update "Penguin 2.0" in the 22nd calendar week

namely the preservation of the "greatest common denominator" as an optimal basis for comparison with last year's analysis, and the consideration of new keywords that have emerged due to growth in search volume in the top 10,000.

The data pool at Searchmetrics is always kept up-to-date. Therefore, new, relevant keywords are used in current studies, such as, for example "Samsung Galaxy S4" or "iPhone 5," which did not even exist at the time of the 2012 study.

Interesting background facts

The analysis of 10,000 keywords has led to:

- 30,000 SERPs (search engine result pages)
- 300,000 titles, descriptions, URLs, etc.
- Approx. 150 GB of data
- Approx. 600,000 AdSense blocks
- Approx. 5,300,000,000 backlinks
- Approx. 4,150,000,000 Facebook shares
- Approx. 12,950,000,000 Facebook likes
- Approx. 600,000,000 Facebook comments
- Approx. 1,000,000,000 tweets
- Approx. 330,000,000 Google plus ones
- Approx. 14,500,000 Pins

The correlation coefficients at a glance

As was already the case with the previous analyses on the subject of "Ranking Factors", the presentation of the correlation between Google's search results and the various influencing factors in this white paper is also based on Spearman correlation. In comparison to 2012, significantly more factors have been taken into account – the most important are shown in the following bar chart. Further factors will be discussed in the respective chapters. The cluster color coding introduced in the previous year was used again:

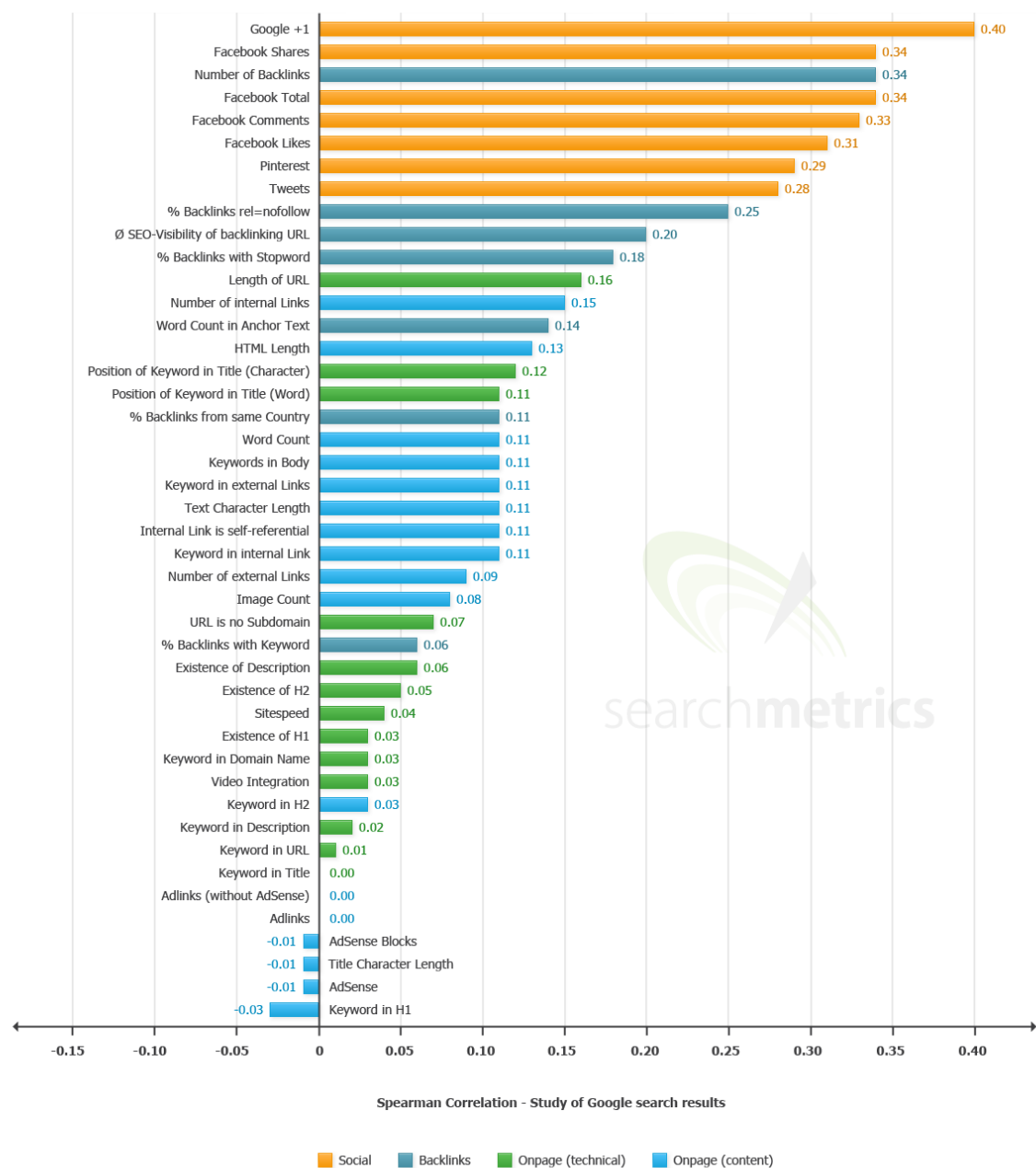


Figure 1: Searchmetrics – SEO Ranking Factors – Rank Correlation 2013

Legend: The correlation coefficient is plotted on the x-axis, which measures its size. The longer/shorter the bar in the positive part of the x-axis, the higher/lower the correlation between that particular factor and a good Google ranking. Our analysis has shown that factors with a correlation factor of zero do not have any measurable correlation with Google's results. Values in the negative range can be positively interpreted with the opposite statement.

1. ON-PAGE CODING

The entire on-page area absolutely belongs to the basics, especially with respect to the technical parameters of a website and the factors affecting a good search engine ranking. It has been the case for some time now *that* fulfilling certain on-page criteria does not have (any longer) a positive effect on the website's ranking. On the other hand, if a website does *not* meet these criteria, this will turn out to be very bad for its ranking.

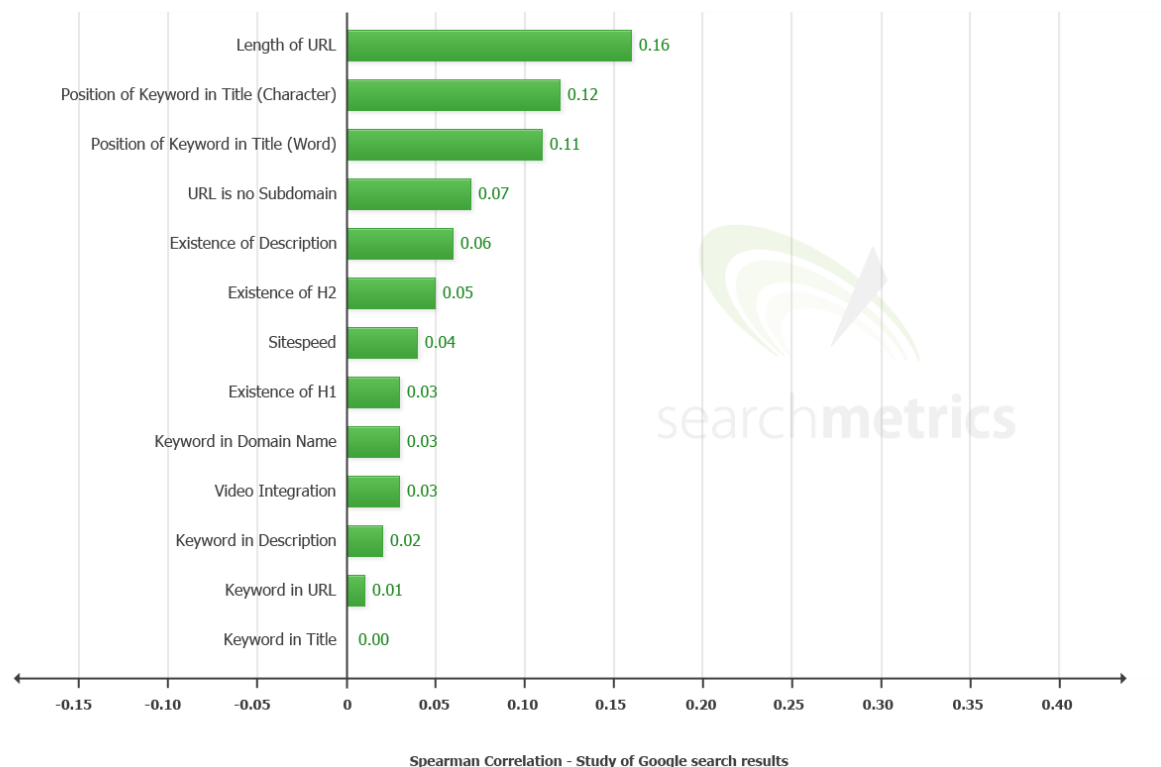


Figure 2: On-page coding – Rank Correlation 2013

Nearly all websites ranked among the top 30 are compatible with the components mentioned later in this chapter. In the case of binary features, in particular, it is hardly imaginable for them to have a positive influence in the presence of factors vs. negative influence in the absence of factors, when in the group of examined elements, the share of elements lacking the analyzed features is so low – i.e.: If almost all analyzed websites meet a specific criterion, a high, positive rank correlation coefficient cannot be logically calculated.

Instead, it appears that on-page coding factors, in particular, represent in many ways the basic prerequisite for a high search result ranking.

In contrast to our results from the previous year, this is no longer related to a domain name or the URL on a keyword basis.

The two newly introduced factors, however, clearly show a certain level of triviality: "URL length" and "URL is not a subdomain". In principle, these two features, which are very closely intertwined, tell us that shorter URLs and domains tend to rank higher on Google. This can be logically explained by the fact that on average across all rankings, the homepage of a domain performs better than individual subdomains or other subpages of this domain.

1.1 On-page coding: Developments since last year

The following detailed view shows the development of the correlation of individual factors since 2012. Features that were included in the study for the first time and for which there is no basis for comparison have been grayed out.

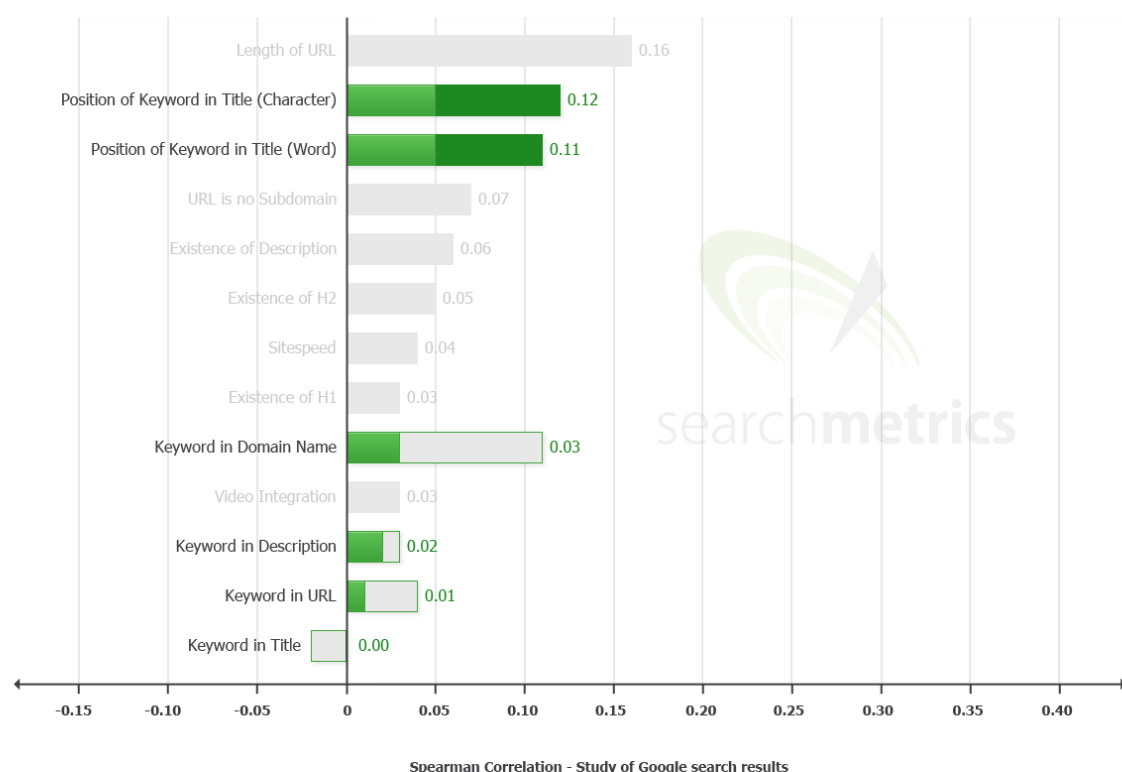


Figure 3: On-page coding 2013 – Changes since 2012

The above chart clearly shows that all but two on-page coding factors display a markedly higher correlation with good rankings than was the case last year. While the relevance of keywords on the page itself, as well as the relevance of the keyword position in the title (the closer to the front, the better) have increased considerably, factors, which used to have a pronounced positive effect on rankings such as keywords in the domain name or the URL, have lost significance.

1.2 Keyword domains are no longer a Ranking Factor

In 2012, the situation was different:

"The power of keyword domains has been known for years and is still clearly visible in our analysis."³

Those days are now over – the once positive effect of keywords in the URL and, in particular, in the domain name has decreased considerably. As a result, the quality of the keyword in the URL/domain as a Ranking Factor is no longer relevant. The following chart clearly shows the sharp drop in correlations for these two features within one year.

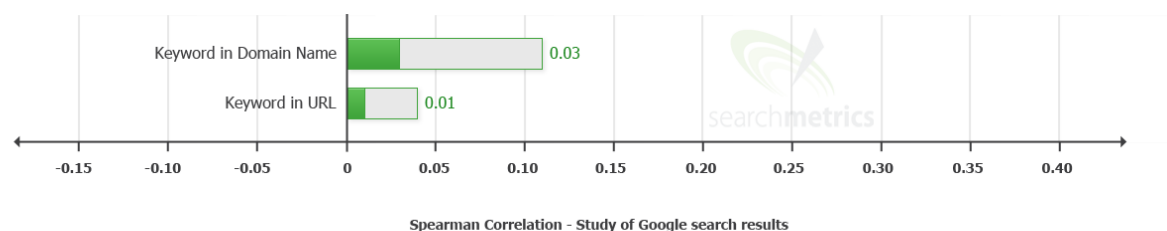


Figure 4: Correlation – Keyword in Domain / URL

The summary chart, which compares the Ranking Factors in 2012 vs. 2013⁴ also shows that these two factors are the biggest losers emerging from this analysis in comparison to the previous year.


This is mainly due to two algorithm changes by Google that were made during 2012:

³ Google Ranking Factors US 2012, Searchmetrics (Chapter 5)

⁴ See Chart 44

The Penguin update, which Google rolled out first in Europe (DE, UK) in March/April and shortly afterwards in the US, was aimed mainly at combating web spam, i.e. websites with low-quality backlinks⁵, i.e. domains with links from low-quality sources "link farms" / link networks and – in the broadest sense – "unnatural" (forced) link building.

Accordingly, there was a massive decrease in the relevance of "hard" keyword links, i.e. backlinks with nothing but the keyword in the anchor text. As a result, the losers of this web spam update tend to drop more sharply in the rankings on a pure keyword level, which, in turn, affects the whole domain through the presence of main keywords in the domain name. For example:

 04/22/2012 GO

Losers (1 to 97) 7 Weekly +







































Keyword	URL	Pos.	Trend	Δ Traffic Index	Search volume	CPC
payday loans	  www.exampledomain.com	75	-74 ↓	-18,604	131,472	€ 4.09
payday loans online	  www.	68	-67 ↓	-4,416	31,201	€ 8.56
pay day loans	  www.	n/a	-100 ↓	-2,324	16,425	€ 6.44
cash advance	  www.	56	-53 ↓	-1,848	48,175	€ 4.20
payday loan	  www.	63	-61 ↓	-1,152	22,547	€ 5.50
paydayloans	  www.	87	-86 ↓	-848	5,987	€ 5.32
online payday loans	  www.	48	-45 ↓	-496	12,970	€ 8.93
check cashing	  www.	63	-59 ↓	-488	14,275	€ 2.63
payday advance	  www.	49	-46 ↓	-444	11,626	€ 6.73
online payday loan	  www.	40	-39 ↓	-400	2,833	€ 6.02
payday advance loans	  www.	59	-58 ↓	-400	2,824	€ 6.41
payday loan online	  www.	42	-41 ↓	-340	2,401	€ 5.27
pay day loan	  www.	51	-49 ↓	-280	5,484	€ 5.87
installment loans	  www.	67	-61 ↓	-236	10,758	€ 4.11
cash advances	  www.	56	-53 ↓	-180	4,713	€ 4.55
payday cash loan	  www.	41	-40 ↓	-176	1,256	€ 3.43
cash loans	  www.	72	-63 ↓	-172	16,483	€ 5.24
fast payday loan	  www.	68	-67 ↓	-164	1,150	€ 4.36
cash loan	  www.	48	-45 ↓	-160	4,222	€ 3.63

Figure 5: Screenshot – Losers at keyword level for an example domain after Penguin update

In this screenshot from "Keyword rankings: Losers of the week" from the Searchmetrics Suite™, we can clearly see a rather drastic loss of rankings at the keyword level in the "Trend" column, which naturally has an impact on the overall performance of the whole website. The domain, from which the data was

⁵ See Chapter 3, Backlinks

collected, is one of the websites, which is about to recover from the impact of the update with the help of new link strategies.

In 2012, Google released two minor updates to Penguin, which was followed by a major update of the algorithm in May 2013. This was launched as Penguin 2.0 and is the second generation of this "over-optimization penalty". This algorithm change occurred after the initial data collection was completed, which is why most of the data had to be updated again. While the effect on the correlation was minor, there was a shift in several results that are only visible when large amounts of data are analyzed (see Wikipedia). Google has already announced further Penguin updates.

The negative developments of the factors "keyword in domain" and "keyword in URL" are also due to the so-called "exact match domain update" (EMD Update). The EMD Update was rolled out by Google in September 2012 and has currently only been implemented in the US – it has not yet been iterated.

In contrast to the Penguin update, which is mainly link-based and works at a keyword level and rather indirectly via the off-page detour in the form of the backlinking source (or its significance as well as the quality of its links), the EMD Update algorithms, while also running at a keyword level, work directly on-page and are URL-based and thus focus directly on keyword domains.

Interestingly, Google only seems to have rolled out the EMD update in the US. No algorithm change has been recorded outside google.com to date. The reason may be perfectly obvious: In other countries – for example, in Germany and also in the UK or France – the ranking of most of the suspicious domains has already been successfully degraded in the course of the Penguin update, and therefore, another algorithm change may not appear necessary in this case.

However, one question still remains: Why is the correlation for the factor "keyword in domain" still positive? Here, too, the answer is not far away: brands.

The special role of brands in the Google ranking already stood out in last year's study. Search engines do not seem to apply the same standards to brands - and their websites - as they do to other domains. The brand factor will play a bigger role in the course of this study.

The fact is that brands – or the related websites – are often quite naturally linked with the name of the brand in the link text alone. What is negative for non-brands, is simply normal for brands. Google is aware of this and treats brands accordingly by not rating this factor as negative in the case of well-known brands, which then rank in the top positions. As a result, there are domains for which the feature "keyword in domain" appears to be positive based on rankings (although other factors, i.e. brand factors – play a key role here), which is the reason behind the still marginally positive correlation of these features.

The following section discusses several factors, which could prove that it is not particularly easy for Google to rank the field behind the brands as the tried and tested on-page methods appear to play again a more significant role here.

1.3 Presence of keywords in the description, title and H1 remains important

Even if the "keyword" factor – as will also be shown later – does have a considerable negative effect in the case of over-optimization on many levels, there are some on-page areas, in which the presence of keywords does not show an above-average correlation with good rankings – because almost all top 30 websites – except for a few brands – have these features. In terms of on-page optimization, keywords remain very important.

The mere presence of a description, H1 and H2 is also important – irrespective of the keyword. These three factors even show a higher correlation with better rankings.

For reasons of clarity, the content factor discussed in the next chapter is included in the following chart, which focuses on these areas.

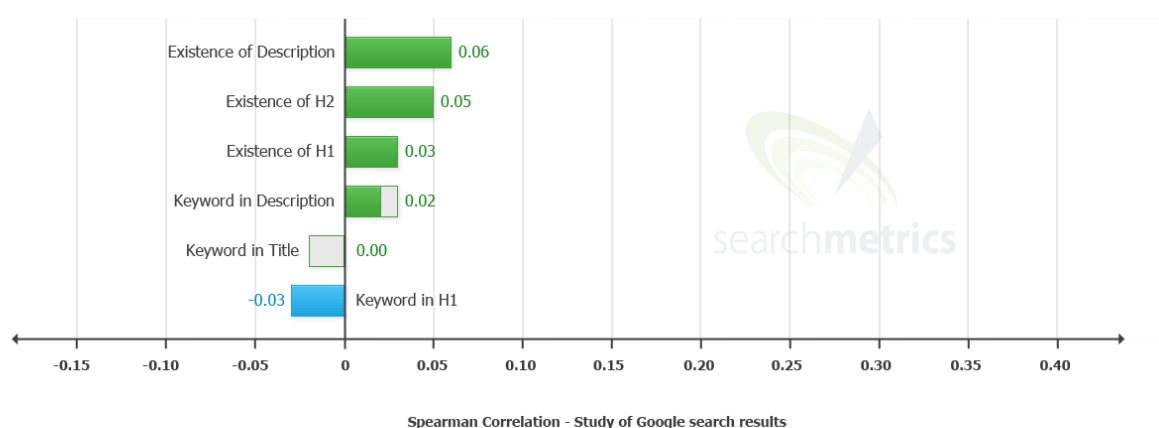


Figure 6: Correlations – Existence of on-page keywords / description, H1 and H2

One cannot deny this development's certain logic. Google is a machine that categorizes data by its relevance to specific search queries. While Google is well on its way to its ideal "semantic search", the process of analyzing content (i.e. words/text) according to semantics (i.e. meaning) probably represents the highest achievement in decoding.

The days of "keyword density" are clearly gone. However, with the exception of the brand factor (which once again has a role in this context), it is – and it has always been – worthwhile to incorporate keywords on-page in the description, the title – and also in the H1 title. These factors are comparable with features of a book, for example. If the title and the description are not attractive, no one will

read it. It is difficult to imagine a book without a description, chapters or headings.

The fact that not only its mere existence, but also the position of a keyword seems to play a role in these lateral areas will be discussed in the section after the digression on the relevance of the so-called $WDF \cdot P \cdot IDF$ formula.

1.4 $WDF \cdot P \cdot IDF$

As already mentioned, "keyword density" as an isolated factor has lost its relevance. In recent times, however, the influence of the so-called $WDF \cdot P \cdot IDF$ formula has been often discussed, which describes the principle of similarity of a query to the page content.

This approach focuses on the correlation of two factors, namely the "number of instances of a keyword in a document in relation to all other subjects "within-document-frequency" (=WDF) and the "number of instances of a keyword in all other relevant documents in the underlying database or inverse-document-frequency (in this case: the Google index)" (=IDF)⁶.

In short: The relevance of a term to a document, and in the second stage, the relevance of a term to a search query using a keyword, should, according to this approach, be weighted with this term/keyword in the index not only in relation to the individual website, but more importantly in relation to all other websites.

The correlation for this factor was also calculated using our data pool for the top 30. Only the content of the website has been taken into account in this case. On the y-axis of the following chart, the average similarity of keywords related to the document in the respective search result ranking is presented on a scale from 0 = "least similar" to 1 = "identical."

⁶ <http://www.karlkratz.de/onlinemarketing-blog/seo-keyword-density/>

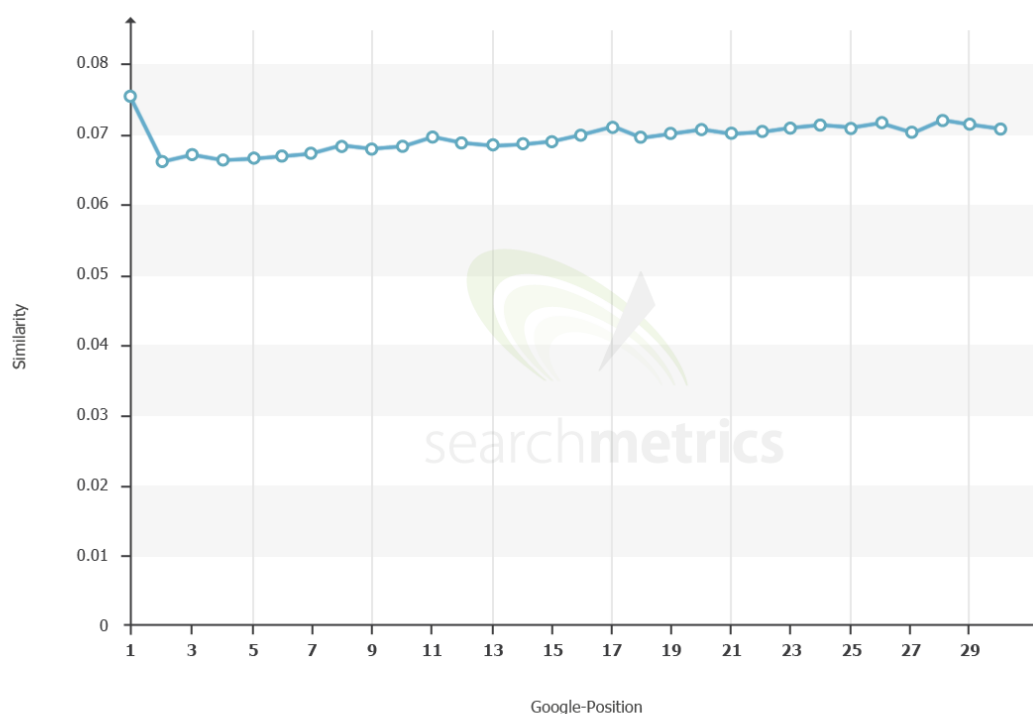


Figure 7: Average by ranking – Similarity

There is almost no correlation – the average is 0.07. On average, the results ranked in the first position seem to exhibit a greater similarity to the search query.

This result suggests that taking the $WDF \cdot P \cdot IDF$ formula for a term, for which an optimized website is already ranked, into account does not have a positive impact on the ranking. Even if the URLs are included in the search results, the ranking for that specific term will not be improved on that basis alone. Instead, the website will more likely improve its search engine performance thanks to the additional use of synonyms and subject-related, relevant terms by getting ranked for other keywords as well.

The relevance of a document in relation to the search query as well as the relevance of all other documents to this term, therefore, does not correlate positively with better rankings, because this factor also appears to be among the trivial prerequisites for a website to be ranked.

In addition to term frequency, keyword proximity and semantic indexing is more important for search engines.

1.5 Keywords in the title should be placed as close to the front as possible

The fact that the presence of keywords in the website title as well as in the description shows good correlation, which has even increased compared with the previous year, has already been proven. Moreover, the position of the keywords in the title itself once again seems to play a greater role in ranking.

The following chart shows the Spearman correlations for the values "position of the keyword" and in relation to the position of the word as well as the characters – according to the formula the values are to be interpreted as *the closer to the front, the better*:

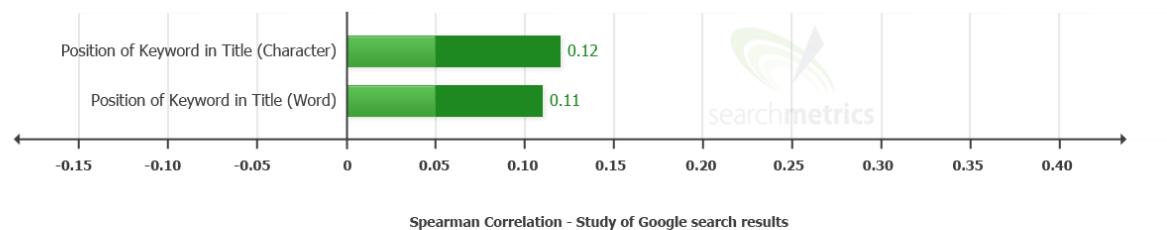


Figure 8: Correlations – Position of the keyword in the title

The positive value for both correlations, which are also closely related, has almost doubled since last year. Accordingly, many websites at the top of the search results have keywords placed as close to the front of the title as possible.

1.6 Page load times as an important performance factor

"Site speed" is a highly-technical feature which was not included in last year's analysis. This factor was included in the pool of correlation factors for 2013 in the expectation that it would show a very positive correlation – but as the summary chart shows, the correlation is only marginally positive.

This may be related to the fact that fast page load times appear to be an utterly trivial precondition of achieving decent rankings on Google similar to the on-page phenomenon mentioned at the beginning of this chapter.

The truth of the matter is that Google measures site speed both directly – when crawling websites – and indirectly – via user signals such as CTR⁷ and bounce rates. Therefore, it is highly likely also a ranking criterion for Google.

The detailed view clearly shows that all the top 30 websites have, in fact, very fast load times and the websites that rank in the first positions load by up to one-tenth of a second faster on average.

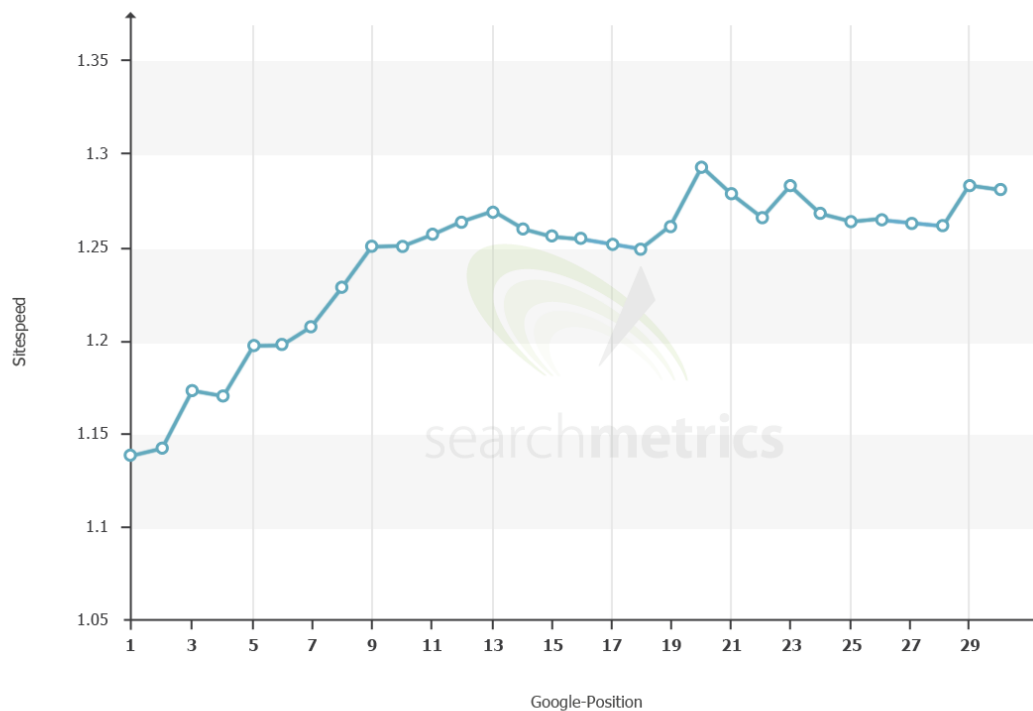


Figure 9: Average by ranking – Page load time

In this context, it should be noted that the factor "site speed" alone only exhibits a marginally positive correlation with good rankings. This is likely due to the fact that the top 30 websites tend to have comparably fast page load times.

1.7 DIGRESSION: Authorship integration does not (yet) have an effect

A technical measure, which appears to gain in importance both on-page and off-page – namely on the social-media level – is authorship integration. Therefore, we will start this section with a short digression on this subject.

⁷ Click Through Rate

Google has made it possible for quite some time to connect URLs with user profiles on Google's own social network Google+ by using the rel="author" tag. This is intended, to assign content in a network clearly to an author, and thus to verify the authorship to some extent. The implementation of this measure by website operators means an immense user assistance in Google's mission to categorize the vast amount of information available on the Internet – and ultimately in the Google index.

Finally, by having authors identify themselves, Google can immediately draw conclusions about the author's popularity and relevance for certain search queries with the help of information about the author received through the verification process from the social chart (connections to other users, activity rate, recognition, status, etc.), as well as relations to other publications and activities of the profile owner in connection with metrics of the website hosting the content itself, as well as user signals from the networks and click behavior from SERPs.

On this basis, the factor "URL has authorship integration" was expected to correlate well with good search result ranking for this URL. This, however, is not the case – at least not to a measurable extent. Other, more in-depth studies on the effect of author integration on rankings conducted by Searchmetrics have so far not delivered any truly positive or negative results.

However, even if there is no direct proof for the positive influence of authorship on the performance of search engines, Google's "authorship integration" will likely become a very relevant factor in the categorization and evaluation of content in specialist areas and niche subjects. When people stand by their content with their name and their reputation, which they may have built over a long period of time, this represents one of the most trusted factors of content verification for search engines, which is also very difficult to manipulate. At the same time, the proportion of spam has fallen in SERPs.

The following chart shows the distribution of URLs with author integration in SERPs. The y-axis shows the percentage of the analyzed URLs with authorship integration, which rank at the respective positions (x-axis).

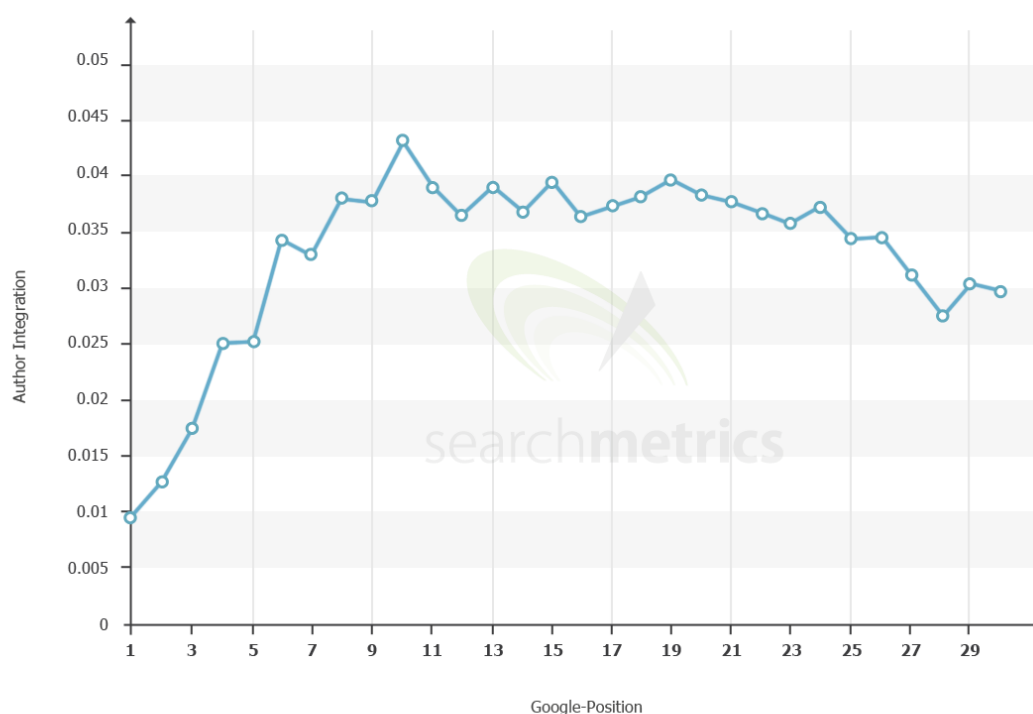


Figure 10: Average by ranking – Presence of authorship integration

As can be seen, most URLs with authorship integration rank on the first search result page further below or on the second SERP. Websites with this feature most often rank approximately between the positions 8 and 20. However, it is much rarer for websites with authorship integration to rank in the top two positions.

In this context, it is, however, important to note that only around 50 percent of analyzed keywords could be included in the calculation of this correlation, which is simply due to the fact that authorship integration is not yet that widespread. However, it is noticeable that snippets accompanied by images with the help of authorship integration achieve significantly better click-through rates than snippets without an author photo – whereby the click-through rates are affected by specific image factors (brightness, background color, attractiveness, etc.).

1.8 Summary

The implementation of on-page coding factors is largely trivial and thus largely a basic prerequisite of achieving top rankings on Google. A well-coded website structure and performance belong to the absolute SEO basics.

With the updates over the past 15 months, Google has sharply discounted the factor "keyword in domain" and thus keyword domains as ranking factors - the on-page presence of the keyword in the title and the description, however,

continues to be very important – as well as the presence of description, H1 and H2. In addition, fast pages load times, as well as the relevance of a document for a keyword in relation to all other documents in the index that are relevant for this keyword, are among the basic requirements for good rankings.

2. ON-PAGE CONTENT

Similar to the coding of their own websites, webmasters can influence the content area, as well. However, these two areas differ in the case of search engines to the extent that technical factors are naturally much easier to crawl for the Google algorithm and thus to categorize⁸.

Content features, however, are often complex and move within a scale with respect to which it has to be determined – from Google's perspective – which value range should be classified as "good" or relevant. After all, it is the content and not the technical implementation that Google wants to evaluate and rate according to its quality and relevance to return a search result.

Even in the case of content features, the number of factors analyzed has increased significantly compared with the previous year.

⁸This is partly due to the fact that most of the factors are binary (presence: yes/no).

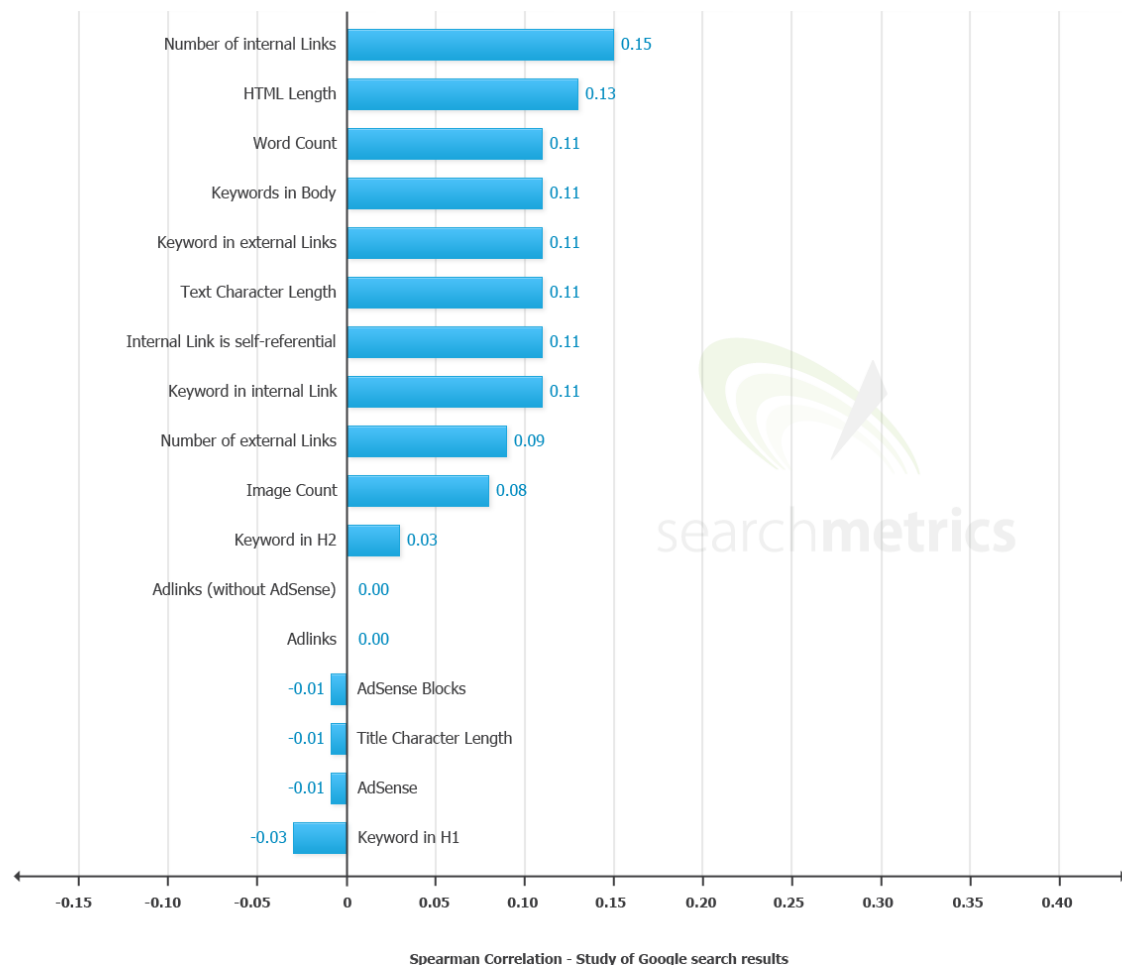


Figure 11: On-page content – Rank Correlation 2013

In principle, the content factors can be broken down into three groups: content quality, linking and advertising.

Except for two values from these areas, all content factors correlate positively with good rankings. In the overview of the year-to-year trends, the increase in relevance of website content becomes apparent.

2.1 On-page content: Developments since last year

The following chart shows the development of the content features that were included in this study in the previous year.

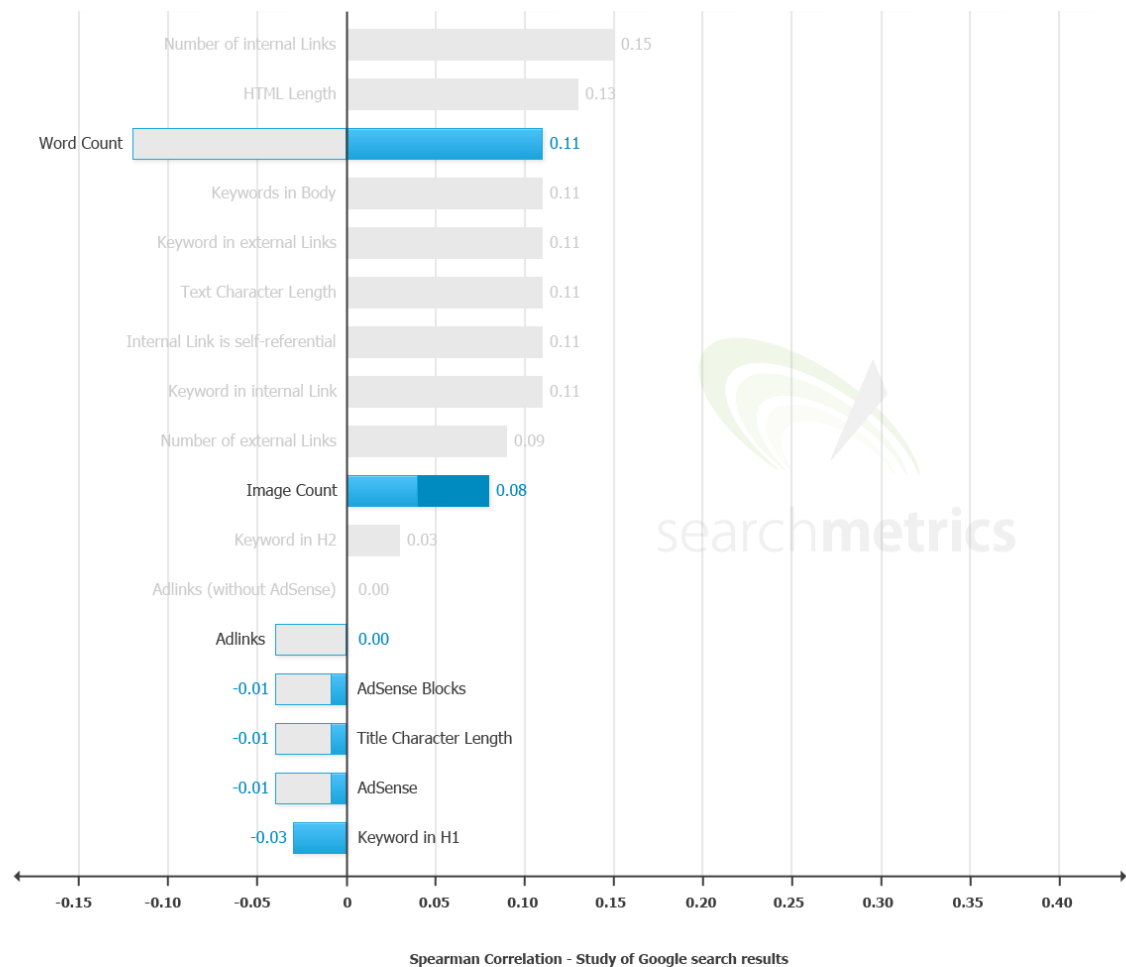


Figure 12: On-page content 2013 – Changes since 2012

It is noticeable that virtually all factors appear to have been upgraded – some quite markedly. Particularly impressive is the development of the factor "word count in text."

The integration of advertising on the website does not seem to be as negative as in the previous year.

The only factor showing a (slightly) negative correlation, next to "keyword in H1" (which is a trivial factor), is "title length." Nevertheless, this can be again interpreted positively: the shorter the page title of a document, the better the ranking.

2.2 Higher number of words in the text is – in contrast to the previous year – clearly important

While in 2012, the feature "word count in text" did still correlate negatively with good rankings, this has changed considerably in 2013.

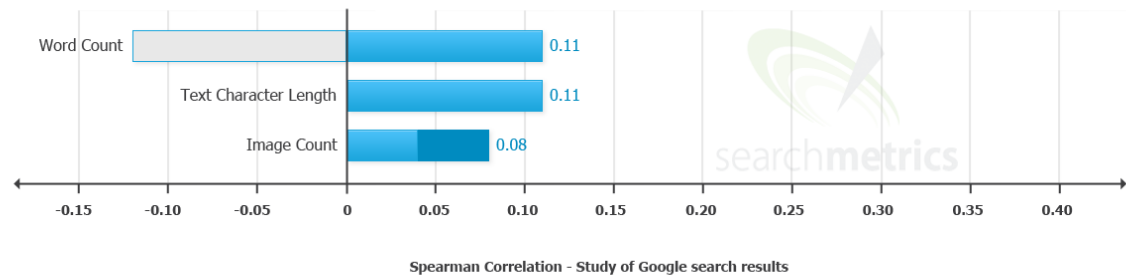


Figure 13: Correlations – Content quality

The detailed chart shows that pages that rank in top positions have a higher word count than websites positioned at the lower end of SERPs. The y-axis shows the average word count for websites for the respective ranks on Google (x-axis).

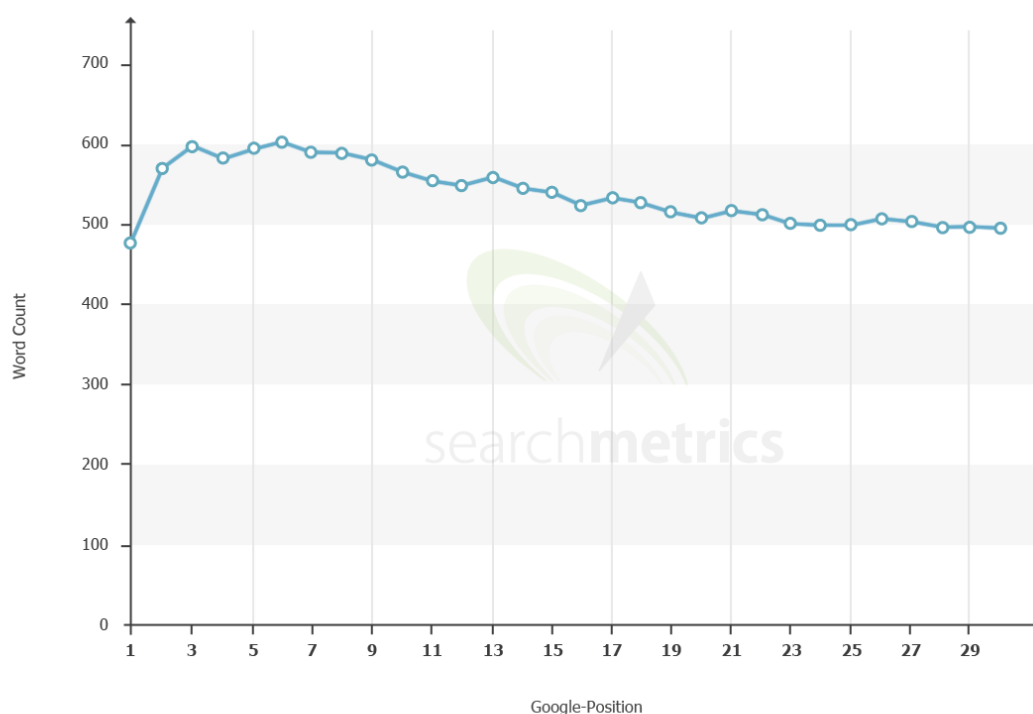


Figure 14: Average by ranking – Word count in text

The presence of something described in this study as "brand factor" is clearly noticeable in the first two positions. The brand factor, which is present in virtually all rank correlation coefficients, will be subject to closer scrutiny in Chapter 5.

In a "normal case" one would expect a continuously falling curve from position 1 (highest value) down to position 30 (lowest value). However, this is not the case because of high ranking of brand websites, which, as already mentioned, occupy a special position. Brand websites seem to have lower word counts on average than search results on the first SERP. Consequently, pages ranked 1st have with 476 words on average 100 words less than pages ranked 3rd (599 words).

2.3 HTML length is important but only to a certain extent

While the length of the HTML code of a website is a technical factor in some respects, the factor "length" in this context depends to a large extent on the type of content, and it will therefore be examined in more detail again in this section.

The majority of websites ranked among the top 30 SERPs tend to have a certain – and not inconsiderable – (minimum) length of HTML code. On average, the analysis of 300,000 HTML documents has resulted in the following correlation of HTML length and ranking (y-axis: average length of HTML code in characters):

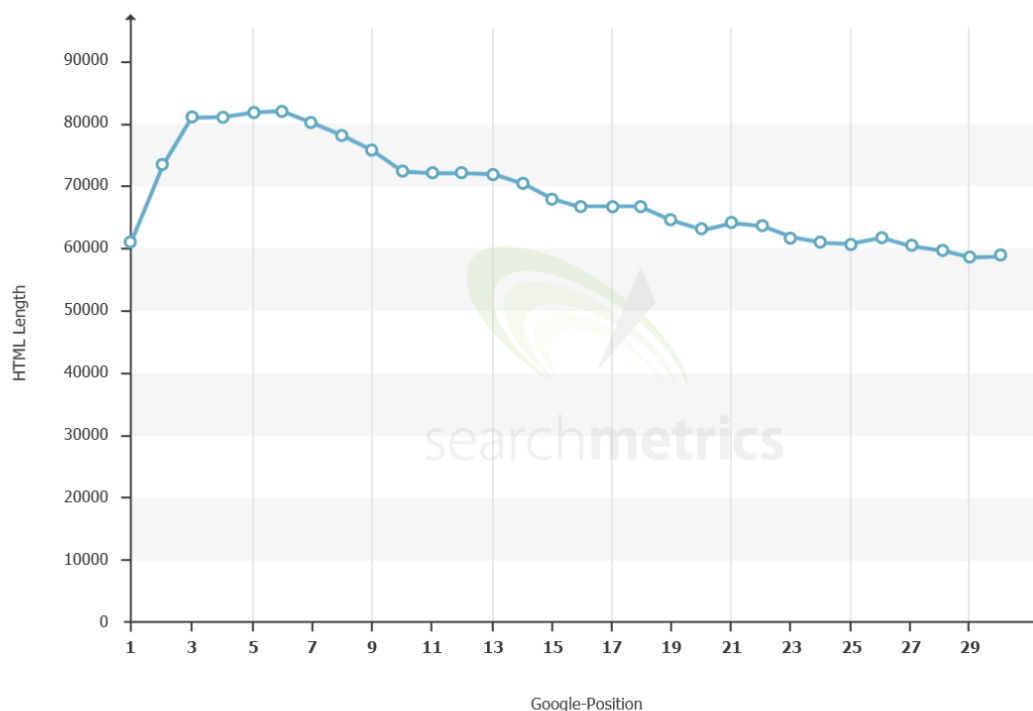


Figure 15: Average by ranking – HTML code length in characters

Again, the brand factor figures into the top ranks. We can observe the curve starting to drop beginning at position 3, which suggests that better ranked documents usually have a longer HTML code than pages in lower positions.

Looking at this chart, it could be assumed that the more characters the HTML code has, the better the ranking. However, this is not the case – at least not without limitation. It should be noted that the optimum is exceeded once a certain limit of approximately 80,000 characters has been reached. While the effect of a bigger number is not necessarily negative, the factor will not improve the ranking results from this point onwards.

2.4 Text length is relevant only up to a certain point

The curve of the coefficient "text length" is nearly identical to the detailed chart of the factor "HTML length". This is logical, because the units are directly interlinked.

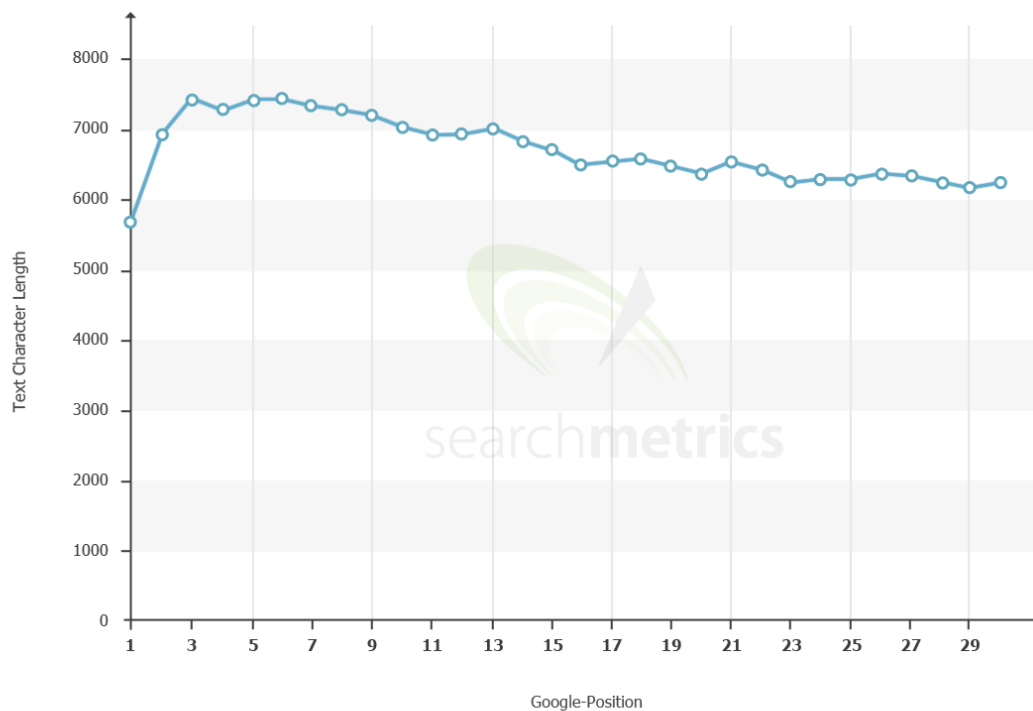


Figure 16: Average by ranking – Text length in characters

Accordingly, the text length is also "capped" with respect to its positive correlation with good rankings. Up to this limit – and excluding the brand factor – one can say, however, that websites with more text achieve a better ranking.

2.5 Using media to enrich content pays off

Images and videos seem not only to make the text more attractive for users, but also for Google. As shown in detail in chart 13, the correlation for the number of images on the website (the more, the better) was already positive in the previous year, and it has increased even further since then.

The following chart shows the entire data pool for image files, i.e. all files with an image tag on the website.

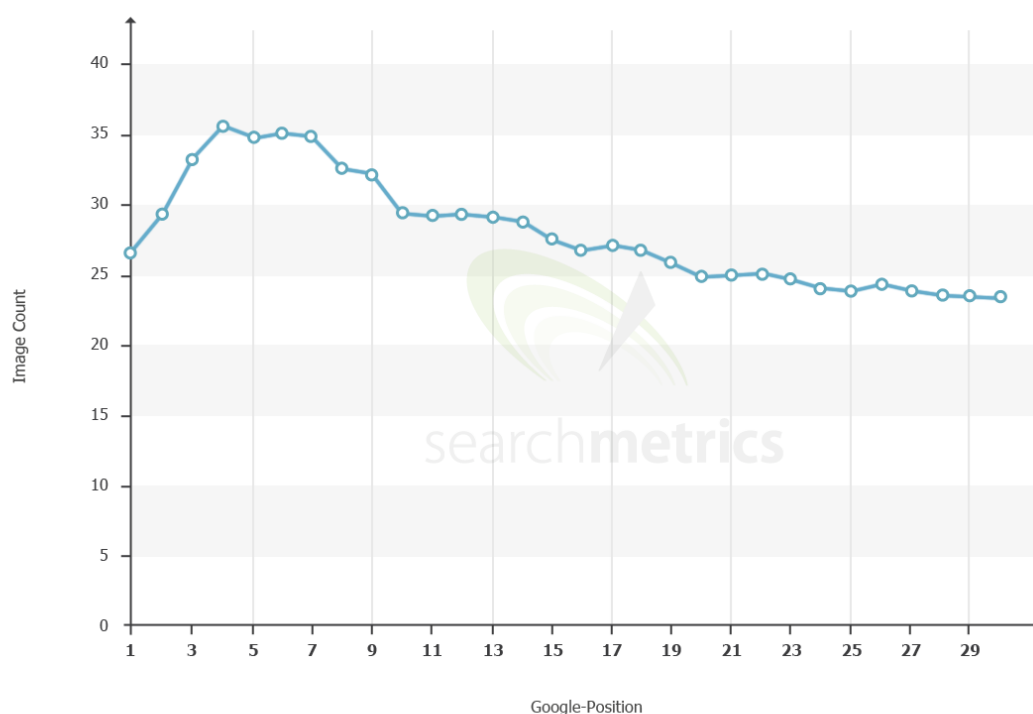


Figure 17: Average by ranking – Image count

Again, it is noticeable that the higher the number of images on a website, the better the ranking. It should be noted in this context that this calculation is not limited to images that are directly embedded in the content. In other words, images in the content as well as image files on the website in general, seem to correlate well with a good website ranking.

2.6 Internal links: Number, structure and the text of links optimize link juice

The internal link structure of a domain name is an important driver behind the search engine performance of the domain. Factors such as the "number of links" and "link text" are crucial for the optimal distribution of so-called *link juice*⁹. However, not only the number of links pointing to the ranked URL is relevant, but also the link structure on the URL itself.

The number of internal links in the content of the analyzed websites is shown below:

⁹ 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.

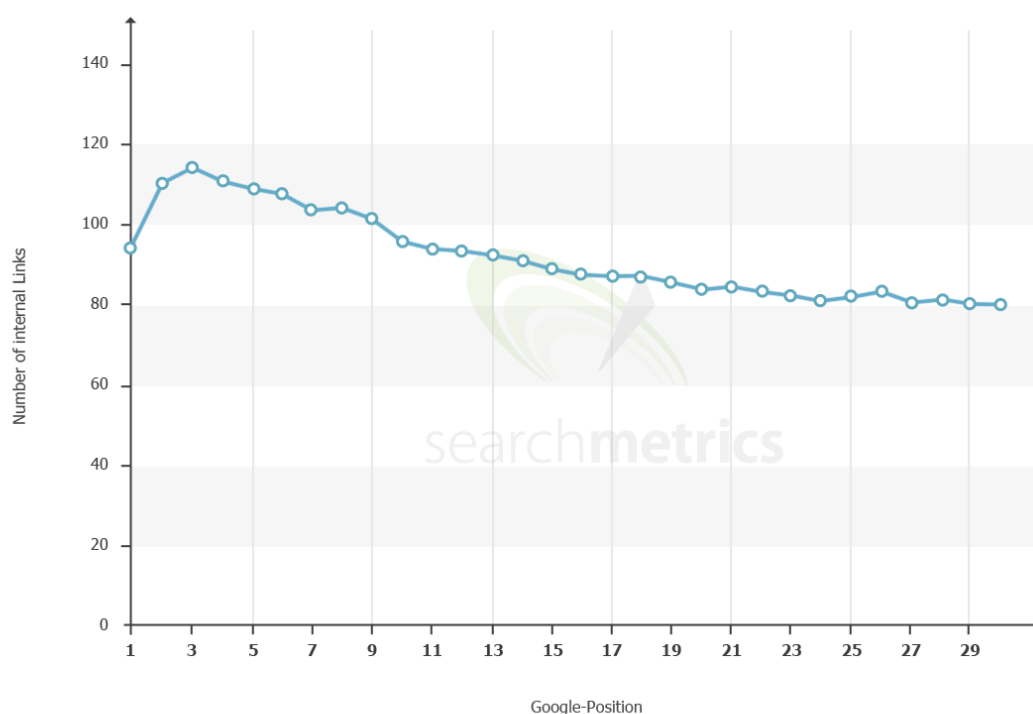


Figure 18: Average by ranking – Number of internal links

The chart clearly shows that the pages in the top search result positions have a comparably greater number of internal links than pages that rank further back. It is also noticeable that brand websites ranked at the very top have a lower number of links than websites ranked in the next positions on average.

The average number of links will naturally increase in the case of websites with strong menu structures and many internal links in the footer. However, not only the number of internal links, but also internally-linked keywords as a link text play a role in the optimal distribution of the link juice. The current rule of thumb suggests to use "hard" links internally (= with keyword/s) and "soft" links externally (= generic / stop words / more words in the anchor, brand links, etc.).

This leads to an interesting question: Does a well-ranked website contain internal links to another website for a keyword for which it is ranked? If this were the case, it would imply, in principle that the website that receives the link would have to be more relevant for that keyword than the website from which the link originates. The ranked website would quasi pass on its link juice and presumably lose its ranking sooner or later.

It is therefore interesting to examine whether websites that perform well use the opposite principle and use the keyword to link to itself. This question is the basis of the analytical approach presented in the following chart:

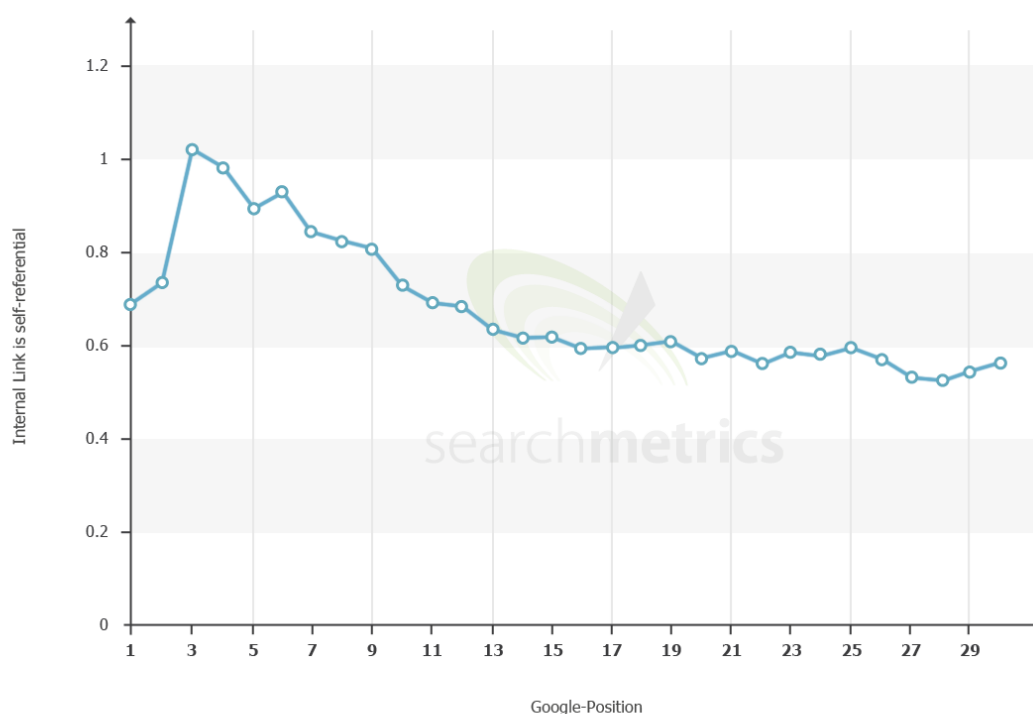


Figure 19: Average by ranking – Number of internal links pointing to the same website

This chart shows the correlation between the average number of self-referential internal links (= links that refer to themselves) of websites, which are plotted on the y-axis, with the respective Google ranking, which is plotted on the x-axis.

From the pool of websites that have at least one internal link, seven out of ten URLs ranked in the 1st place have one link to themselves on average, while websites that rank in the 3rd place contain on average one self-referential link.

This number is naturally influenced by several websites, which have more than one self-referential link. In the case of some domains and associated URLs, this phenomenon is purely technical – for example, on Twitter, where each subpage links back to the profile itself with a tweet in the form of a permanently linked address of the user (@profile name).

When examining the specific results, it becomes apparent that those top-ranking websites (including many shopping websites) that refer to themselves using a keyword tend to do this in the site-wide menu or breadcrumb link.

It should be noted that using a keyword for which one website is supposed to be ranked to establish a link to another website should be avoided. However, many websites that rank in the top positions create another internal link to themselves using the keyword – mainly on a structurally-technical basis.

2.7 Moderate integration of AdSense and other Adlinks is less negative than before

In 2012 there were clearly negative correlations with good rankings with respect to the integration of advertising –even for AdSense, Google's own product. This means that last year, well-ranked websites had less advertising than those ranked lower. In 2013, these correlations are close to neutral.

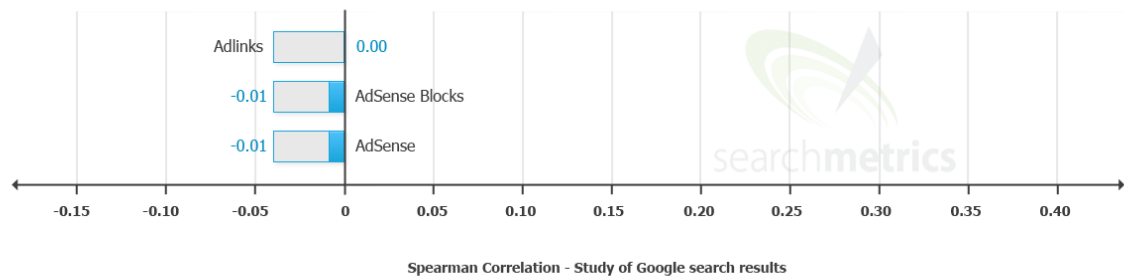


Figure 20: Correlations – Ad integration

This chart shows that even well-ranked websites are now integrating advertising, and the number of ads is rising. The following charts illustrate this in greater detail.

Legend: The y-axis of the charts 20 and 21 is scaled from 0 to 1, and multiplied by 100 it returns the percentage of websites that rank in a particular position for which the respective factor applies (y-axis caption).

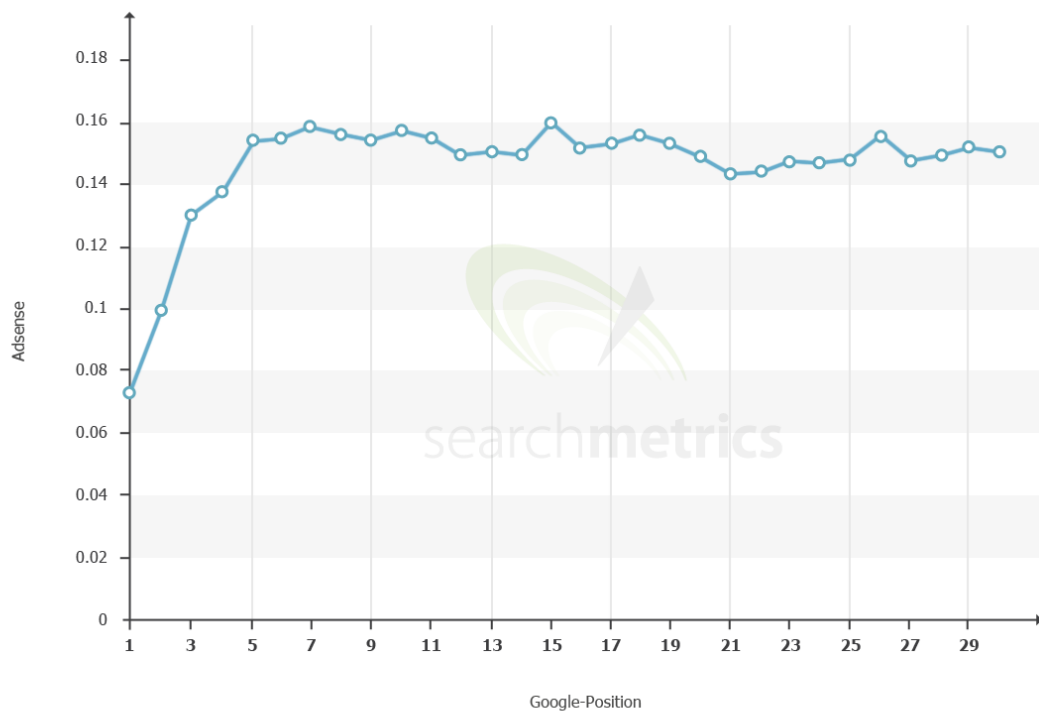


Figure 21: Average by ranking – AdSense

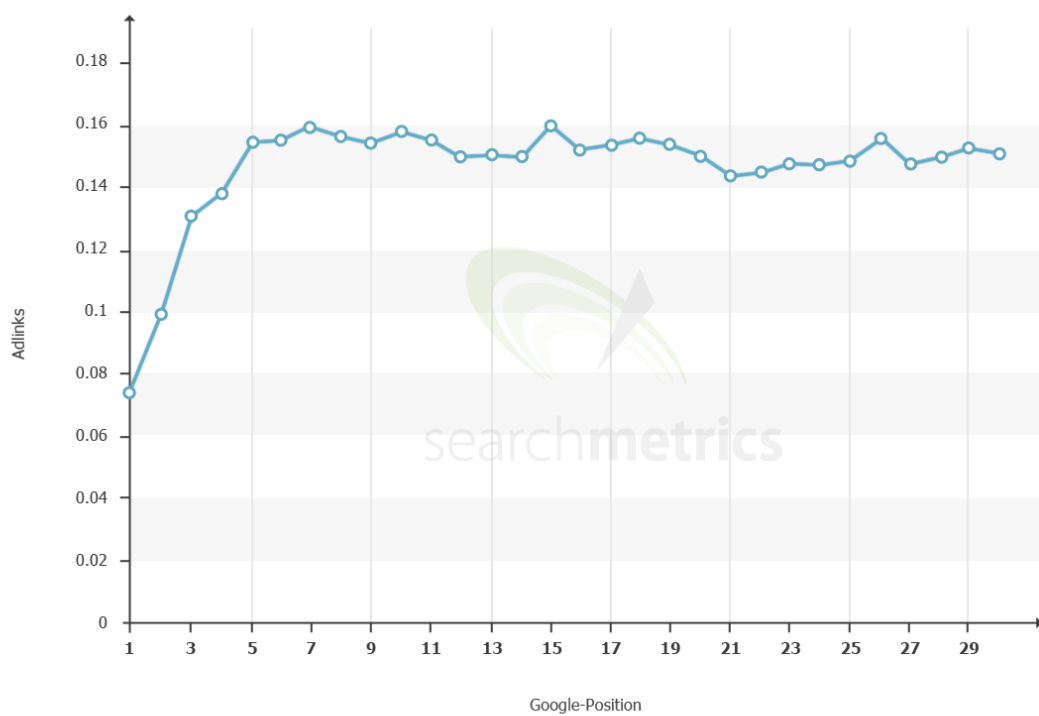


Figure 22: Average by ranking – Adlinks total (including AdSense)

The main message of these two charts is that especially among the top 5 URLs in terms of search results, there are on average fewer pages with at least one ad integration (AdSense or another form of Adlinks) than is the case with URLs ranked in the following places. Lower ranking tend to have more advertising than the the top 4.

Incidentally, the results for charts 21/22 are not identical. This is simply due to the fact that the vast majority of websites use AdSense rather than another form of Adlink. According to this, most of the top 30 websites use the advertising integration from Google itself – in fact, the integration of other Adlinks such as AdSense shows a slightly higher correlation with good rankings than in the previous year.

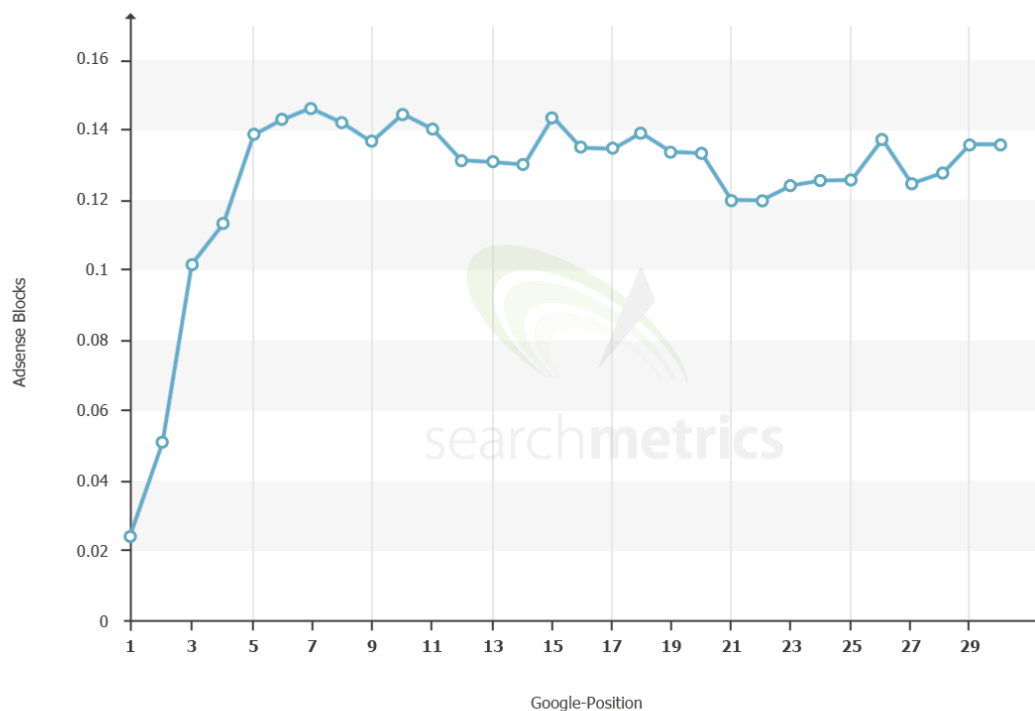


Figure 23: Average by ranking – Number of AdSense ads

Even in relation to the number of AdSense blocks, well-ranked websites are positioned below the rest of the top 30. The top five ranked URLs in terms of search results feature less advertising integration than URLs in the second half of the first SERP. From the second SERP on, the number of AdSense integrations remains nearly unchanged on average, or it even drops slightly.

2.8 Summary

The quality of content was already very important in the previous year, which has become even more the case in 2013. Well-ranked URLs tend, to a certain extent, to have more text and even more multimedia content than was the case in 2012. In addition, the moderate involvement of advertising – even outside of AdSense – does not seem to have as negative of an effect on rankings than before. While keywords in H1 or H2 titles do not seem to be crucial for rankings, they are almost trivial. Well-correlated feature combinations were not identified.

A good internal link structure is also an important quality factor. It seems to have a positive effect, when the most relevant URL of a domain for a particular keyword uses this keyword once again to create a link to itself. On average, the majority of top-ranking URLs contain one self-referential link.

3. BACKLINKS

The quality factors in the area of backlinks are probably the most important SEO metrics on the off-page level. Google's success is based not least on the criterion of "PageRank", the metrics of which have always been backlink-heavy. Links from other websites, therefore, are historically among the most important indications of relevance in the form of recommendations while taking the main subject of the website into account. Any possibilities of manipulation that existed in the past have been continually minimized by the aforementioned iterative algorithm changes performed by the search engines.

As a result, the ranking of a website is determined not only by the number of backlinks, but also the quality of the backlink structure, which, in turn, is subject to various factors. In order to reflect the diversity and complexity of the backlink cluster, the factor pool for the analysis of backlinks was expanded to include several parameters, including its own metric of SEO visibility.

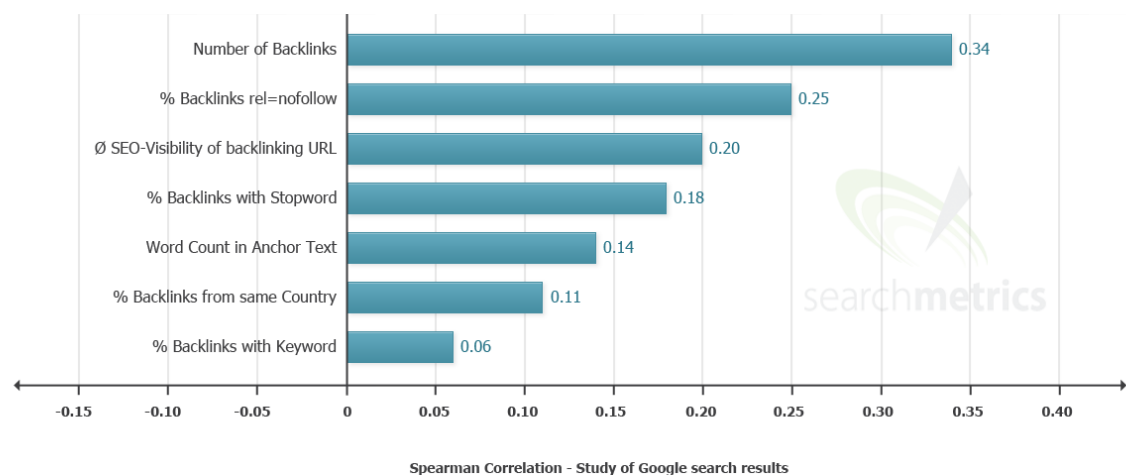


Figure 24: Backlinks – Rank Correlation 2013

The Penguin and the EMD update in the US have had a decisive impact on the development of correlations in the course of the year¹⁰. Matt Cutts, Google's "Head of Webspam" described the importance and the extent of the latest

¹⁰<http://searchengineland.com/penguin-2-0-losers-porn-sites-game-sites-big-brands-like-dish-com-the-salvation-army-160744/>

Penguin update shortly after its roll-out on his blog¹¹. In comparison to the previous year, this factor has changed as follows:

3.1 Backlinks: Developments since last year

The following chart shows the development of four factors compared to the previous year.

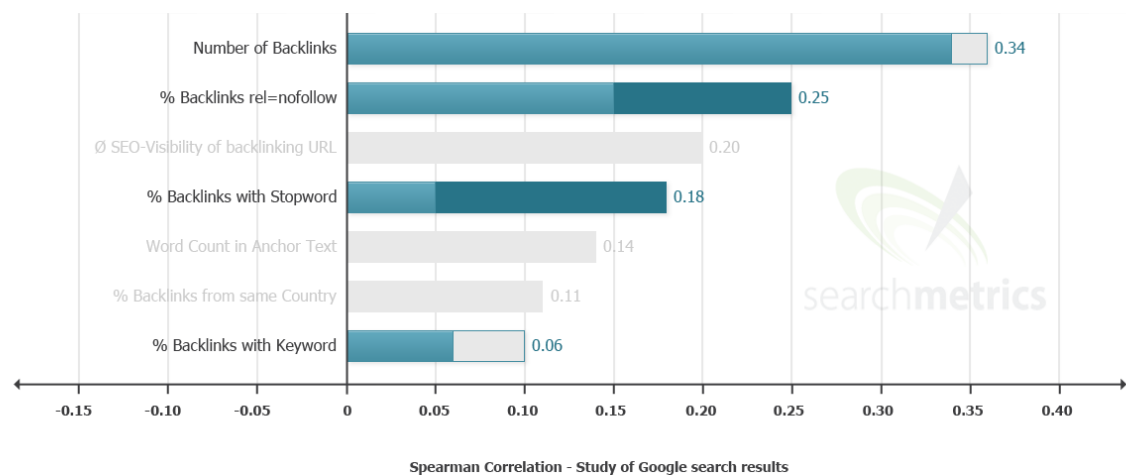


Figure 25: Backlinks 2013 – Changes since 2012

Websites that rank in top search result positions contain an even higher number of backlinks than was the case last year. Moreover, the number of links as shown in the summary chart exhibits one of the highest correlations with good rankings, which becomes the highest once the correlations of the social area are taken out of the equation.

The websites are not only required to have a greater number of links, but the relevance of the quality of links has also increased.

As a result, well-ranked websites tend to have a significantly higher proportion of both "nofollow" links as well as backlinks with a stop word, while the proportion of back links with a keyword has been markedly reduced. The correlations between these quality factors have shifted quite considerably. As anticipated for some time, links that have been "artificially" keyword optimized appear to have been discounted in favor of a "natural" link profile.

¹¹ <http://www.mattcutts.com/blog/penguin-2-0-rolled-out-today/>

3.2 The number of backlinks remains immensely important

The chart shows in detail the massive difference between websites ranked in the first place and further in the back:

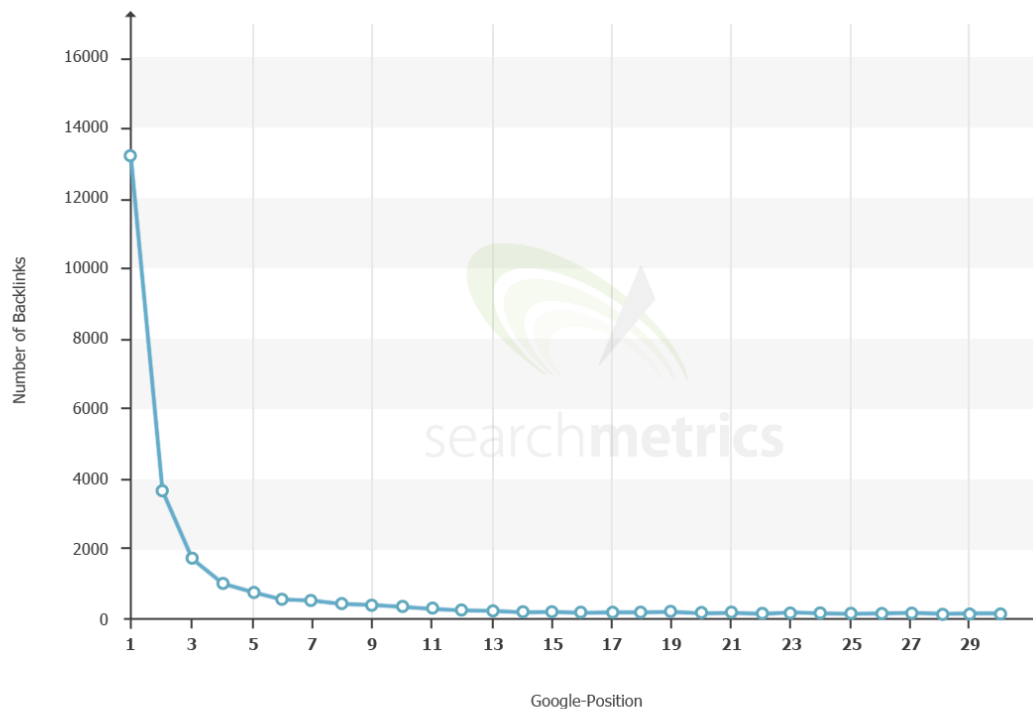


Figure 26: Average by ranking – Number of backlinks

This chart shows the number of links in relation to the ranking of the linked websites. While websites ranked 1st have around 13,358 backlinks¹² on average, this figure drops for websites ranked 2nd to 3,693 backlinks. Websites ranked in the 30th position only have just under 103 backlinks on average.

3.3 Link diversity is always relevant

The positive trends with respect to link quality factors, which have displayed an even higher correlation with good rankings in 2013 than was the case in 2012, suggest that a more differentiated link building strategy is required in order to achieve the goal of optimized rankings.

¹² Excluding Wikipedia URLs. When Wikipedia URLs are included, the average decreases by around 50%

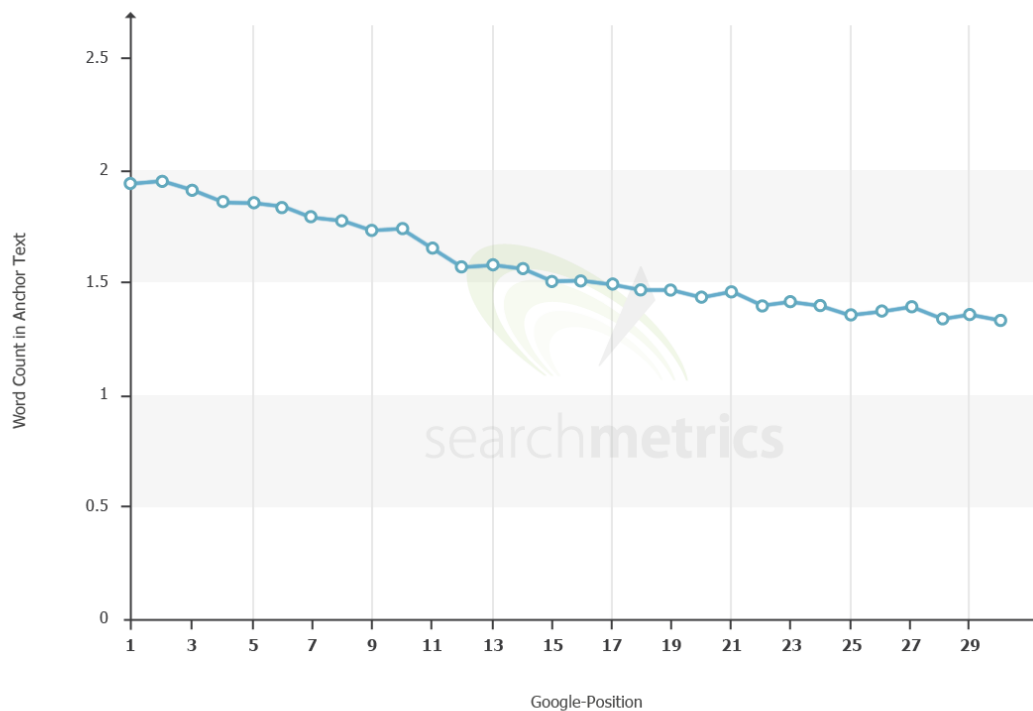


Figure 27: Average by ranking – number of words in anchor text

The "soft" factors on the backlink level seem to be gaining in relevance. Well-ranked websites not only have a significantly higher proportion of multi-word anchor texts in the backlink profile, but also feature "nofollow" links, which is demonstrated in the following chart (y-axis: the proportion in % is obtained by multiplying the result by 100):

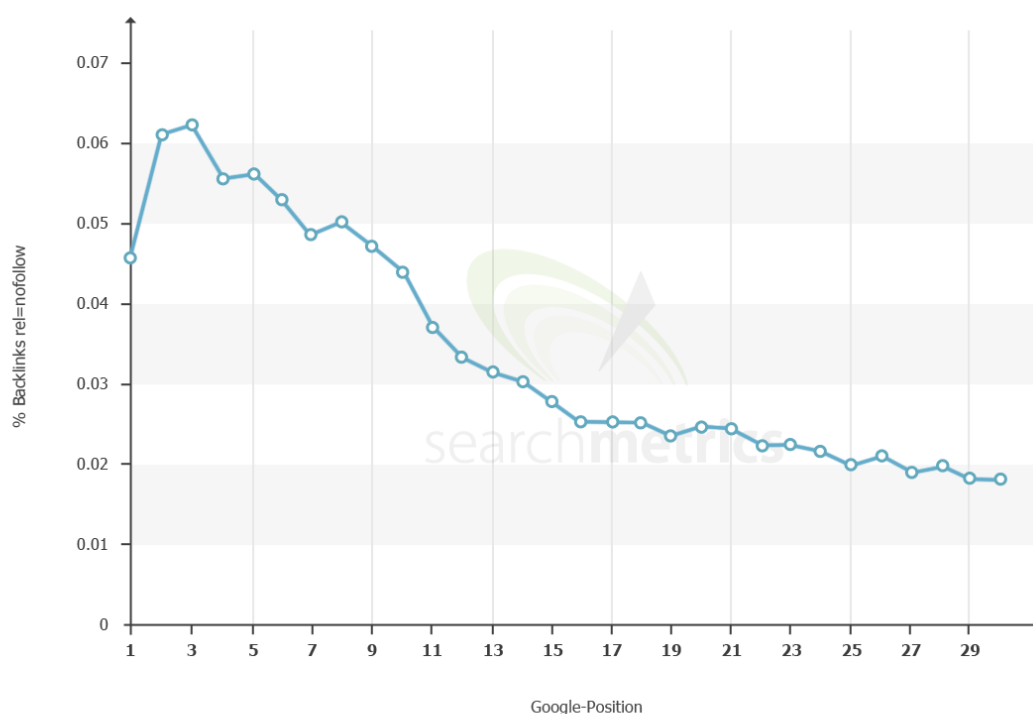


Figure 28: Average by ranking – Percentage of nofollow backlinks

The relevant correlations with link diversity are clearly pronounced, and the upward trend is expected to continue. In conclusion, this means that the factor "keyword in anchor text" – or the "keyword as anchor text" – has lost its relevance. Indeed, a look at the correlation of this factor shows a clear result.

3.4 "Hard" keyword anchor text has lost its relevance

The days of "hard" keyword links appear to be numbered. The positive correlation for this factor across the top 30 has decreased significantly during the year and is now barely in the positive category. The downward trend is expected to continue.

The detailed view shows quite clearly that this factor does not (any longer) have a well-formed correlation. The shape of the curve remains virtually unchanged.

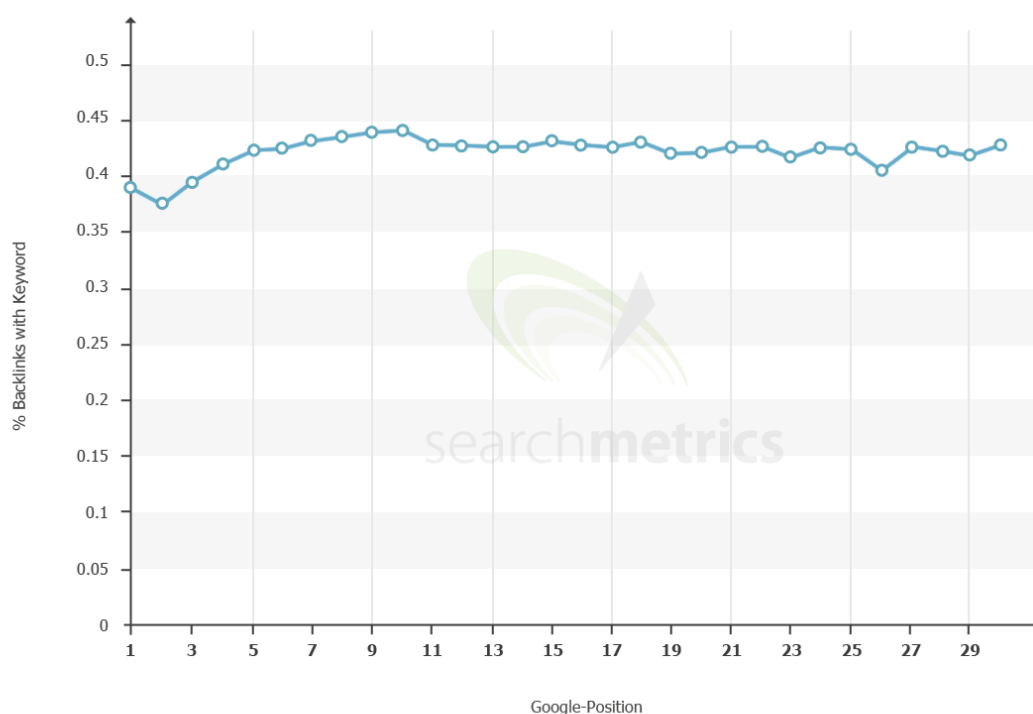


Figure 29: Average by ranking – percentage of backlinks with keyword

Nevertheless, it should be noted that in absolute terms, backlinks with a keyword clearly dominate the backlink portfolio of well-ranked websites (the proportion in % plotted on the y-axis is obtained by multiplying the results by 100). On average 42-43 percent of backlinks of websites ranked among the top 30 contain keywords in the anchor text.

This is not surprising – after all, having keywords in the link text is part of a natural backlink profile – in particular in the form of brand links. Crucial in this context is, however, the distribution of individual link parameters within the backlink profile as well as the trend, which is sharply downward.

This is also noticeable with regard to the proportion of backlinks with stop words, which has actually not only increased significantly for websites ranked at the top of search results but also shows a particularly strong correlation:

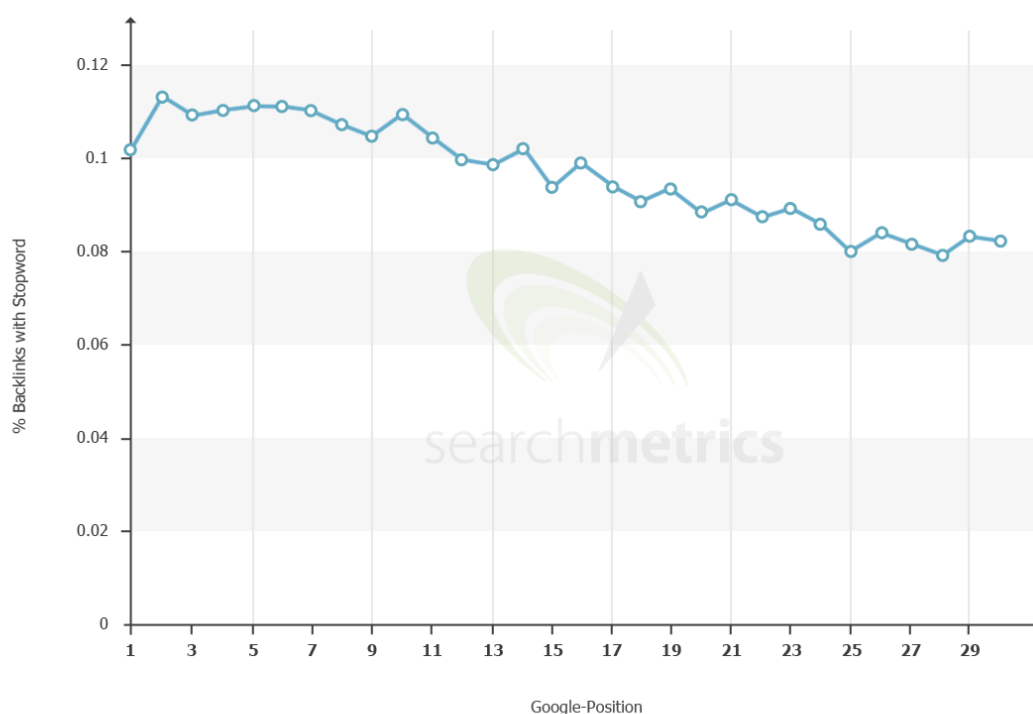


Figure 30: Average by ranking – Percentage of backlinks with stop word

The URLs in the first four places have a higher proportion of links with stop words than websites that rank below them. From Google's point of view, it seems to be simply "unnatural" when individual URLs are always linked with the same keyword or anchor text.

At the very least it is highly unusual that a large number of different webmasters (authors, etc.) would independently link a website with exactly the same anchor text. In any case, Google tends to suspect manipulation – and probably quite justifiably – in the case of link profiles, where the proportion of hard keyword links exceeds a certain tolerance level. For this reason, Google has been depreciating this factor. It is, therefore, clearly more important to increase the number of links – in particular, brand links, links with stop words, "nofollow" links – and at the same time to decrease the number of keyword links.

3.5 Backlinks from URLs with high SEO Visibility are significantly more valuable

In introducing this subsection, it is necessary to explain the term "SEO Visibility" first. This metric, which is also referred to as "visibility" and has its origins in SEO analysis software tools, is calculated from a whole range of various factors related to keyword rankings of a fixed keyword pool and used similarly to factor

traffic (which is not measurable from the outside by non-webmasters) as a basic factor for the quality categorization of websites for comparison purposes. In addition to keyword ranking, Searchmetrics also takes into consideration search volume, information about keyword characteristics (e.g. navigational or informational search) as well as the layout of the search result. In short: The higher the SEO Visibility of a website, the more traffic it likely generates¹³.

The correlation of the factor of "SEO Visibility of the backlinking URL", reveals that the traffic of URL B – in terms of its relevance as a source for a link to URL A – is crucial for the ranking of URL A.

In short: Backlinks from websites with high SEO Visibility seem to be much more valuable than links from websites with low SEO Visibility.

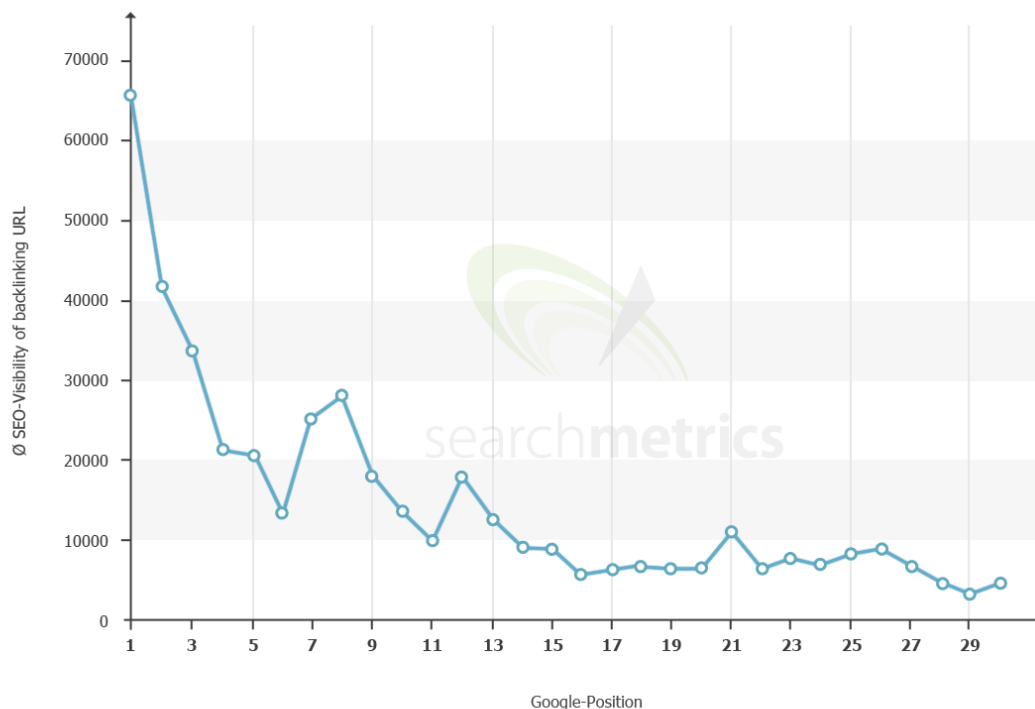


Figure 31: Average by ranking – Average SEO visibility of the link source

Again, top ranking URLs clearly stand out with their backlink profile from other URLs that rank further down. While websites ranked in the 1st place tend to have links from URLs with an average SEO Vvisibility of over 65,000¹⁴ – in the 2nd place this drops to an average visibility of around 42,000 and in the 30th place to only around 4,300.

¹³ The SEO Visibility of domains is calculated for 20 countries and is available for free at <http://suite.searchmetrics.com/us/research>

¹⁴Excluding Wikipedia URLs. With Wikipedia URLs for pos. 1: around 20,000 SEO Visibility

3.6 Combinations with backlink features

Combined with other factors, backlink factors have with other factors from this study a very positive influence – the most significant from all areas. This means that when a feature from the feature pool being analyzed here emerges together with factors from the backlink area, the combination of these two features shows a significantly higher correlation.

This means that a large number of top-ranking websites feature backlink factors analyzed in this study, which show positive correlation to such an extent that their value will increase when combined with other coefficients. This clearly shows the importance of backlinks and their direct correlation with good rankings.

For example, the combination of "number of backlinks" and "Facebook shares" with a value of 0.37 exhibits a very high correlation with good rankings. The number of words in the anchor text in combination with the feature "keyword in description" (0.13) also seems to have a positive impact.

In addition to SEO Visibility, which correlates well with "tweets" (0.31) and the attribute "nofollow link" (0.27), nationality¹⁵, also appears to play an important role in this context. Having a large number of backlinks from domains from the same country combined with "Facebook likes" (0.32), "Number of words" (0.14) or "tweets" (0.31), for example, also has a positive effect on the site's ranking.

3.7 Summary

Backlinks continue to be one of the most critical SEO metrics. Over the years, very little has changed to diminish the importance of the factor "number of backlinks". Websites with more links rank higher in search engines. However, the factors related to this metric have evolved over time. As a result, it is no longer only the quantity of backlinks which is crucial, but also increasingly their quality.

Google has significantly discounted several keyword factors – especially during the last year. Naturally, a certain proportion – and not necessarily a small one – of so called "hard" anchor text or "keyword links" is still considered to be "natural" to a certain extent. Not least through its Penguin updates and iterations, Google has changed the tolerance limits decisively.

As a result, well-ranked sites feature a balanced mix of "nofollow" links, multi-word anchor texts, link text with a stop word share and keyword links. The quality of the link source also plays a crucial role. Google's algorithms and the group's own quality raters are getting better at unmasking link farms. Links from

¹⁵ In relation to the location of the server

websites with high SEO Visibility are much more valuable. The backlink profile is nothing more than a collection of very complex quality factors.

4. SOCIAL SIGNALS

On average, social signals are the domain with the highest correlations. At the beginning of this chapter, we want to reiterate that correlation is not synonymous with causation.

More detailed analyses of the actual impact of social signals on the indexing and ranking of URLs conducted in isolation from other factors were part of several additional studies, which will be discussed in more detail later in this chapter.

Nonetheless, the following fact remains completely disconnected from causal aspects: Websites that rank in the top positions on Google usually have a large number of social signals. In other words: well-ranked URLs have many shares, likes, comments, plus ones and tweets.

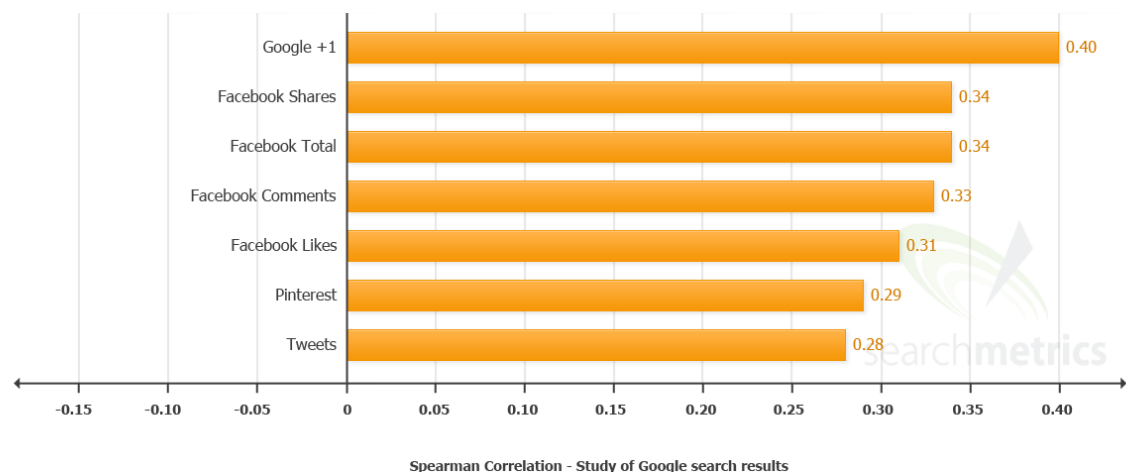


Figure 32: Social signals – Rank Correlation 2013

In addition, the strong correlation of the entire social sector is supported by the fact that the number of social signals seems to drop in line with the ranking of websites in SERPs – the worse the ranking, the lower the number of social signals. This is particularly evident in the detailed charts shown in the following subchapters. However, we will focus our attention first on the development of correlation of social signals since last year.

4.1 Social Signals: Developments since last year

The following chart shows that social signals that already had a very high correlation with good rankings in 2012 feature even more frequently and strongly in URLs well-positioned in the SERPs in 2013.

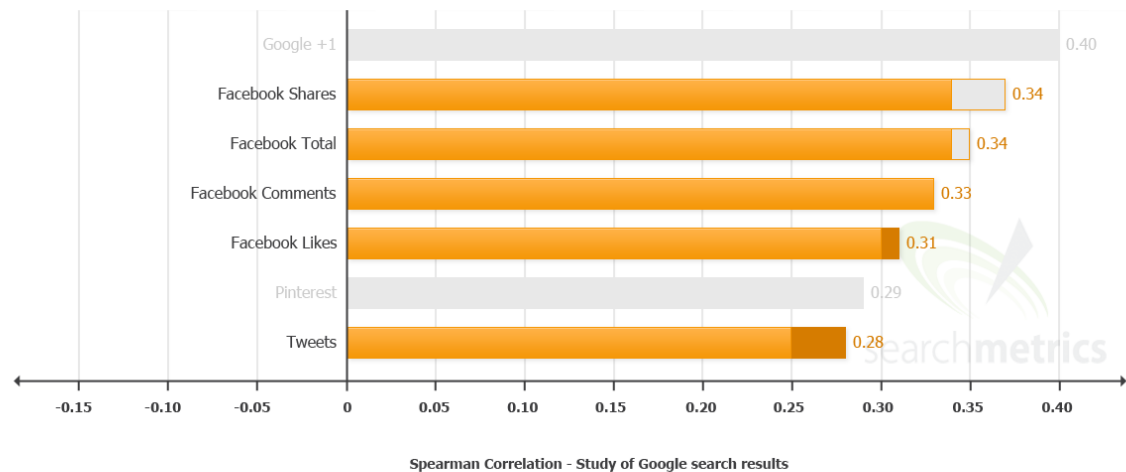


Figure 33: Social signals 2013 – Changes since 2012

Facebook signals as well as tweets are showing slightly higher correlation than in the previous year.

A new addition to the factor pool are signals from Google's own network Google+, which also shows a strong correlation with good rankings, as well as social signals from the aggressively promoted photo community Pinterest.

4.2 Social signals continue to show strong correlation with good rankings

The correlation of social metrics with better rankings is impressively high. This has already been the case last year, and it is true again in 2013. Recommendations and related signals from the social chart have a strong correlation with good rankings on Google.

The detailed charts for individual networks have distinct curve shapes, which are used to calculate the strong correlation.

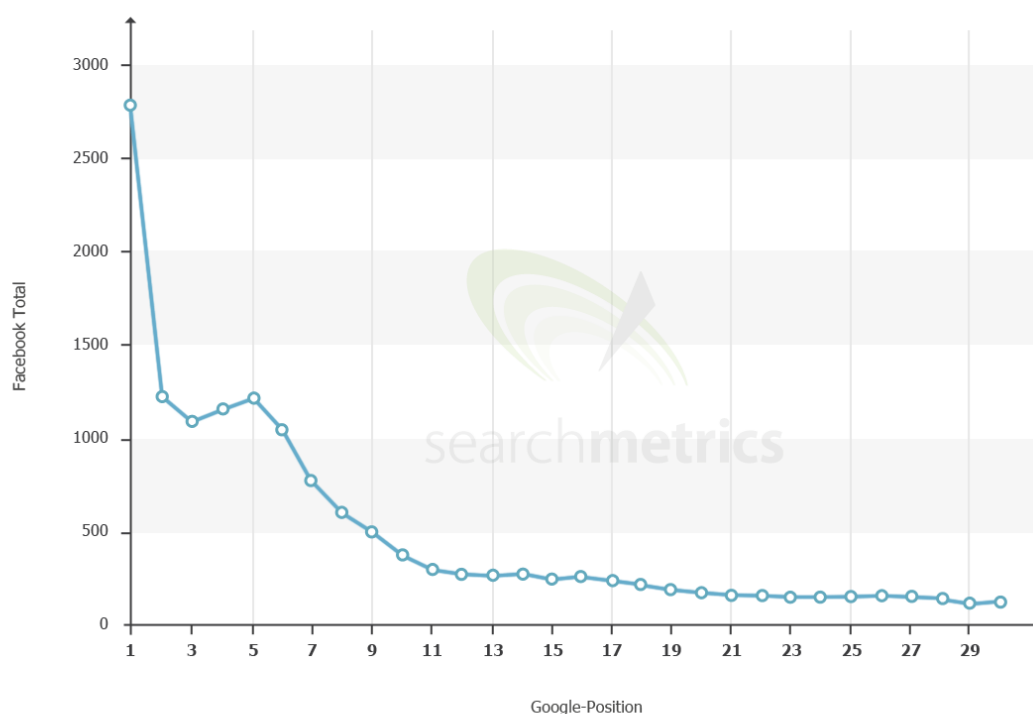


Figure 34: Average by ranking – Facebook signals in total

As an example, we show here a correlation of the summary of all Facebook metrics (shares, likes and comments) for the factor "Facebook total" which has – as an aggregate of all factors – the highest correlation and by far the highest average result on the y-axis when compared to the Google rank (x-axis) within the social area.

The average number of tweets from the Twitter network – which is more of an information aggregator than a social network – is in relation to the Google rank of the relevant URL already well below the figures for Facebook's signals. Even here, the correlation with rankings is clearly visible.

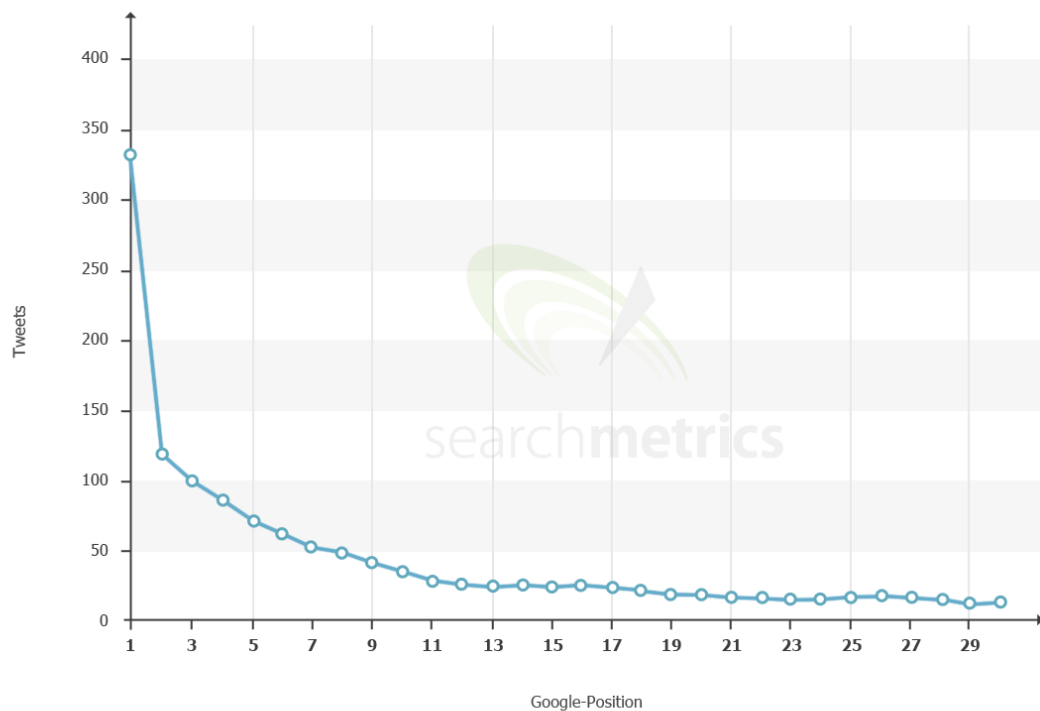


Figure 35: Average by ranking – Twitter signals

Pins – i.e. signals from the Pinterest network, which allows members to "pin" content, mainly in the form of images to their so-called pinboards – are comparatively rare as a social signal per website and ranking. However, even here the distribution reflects a fairly good correlation.

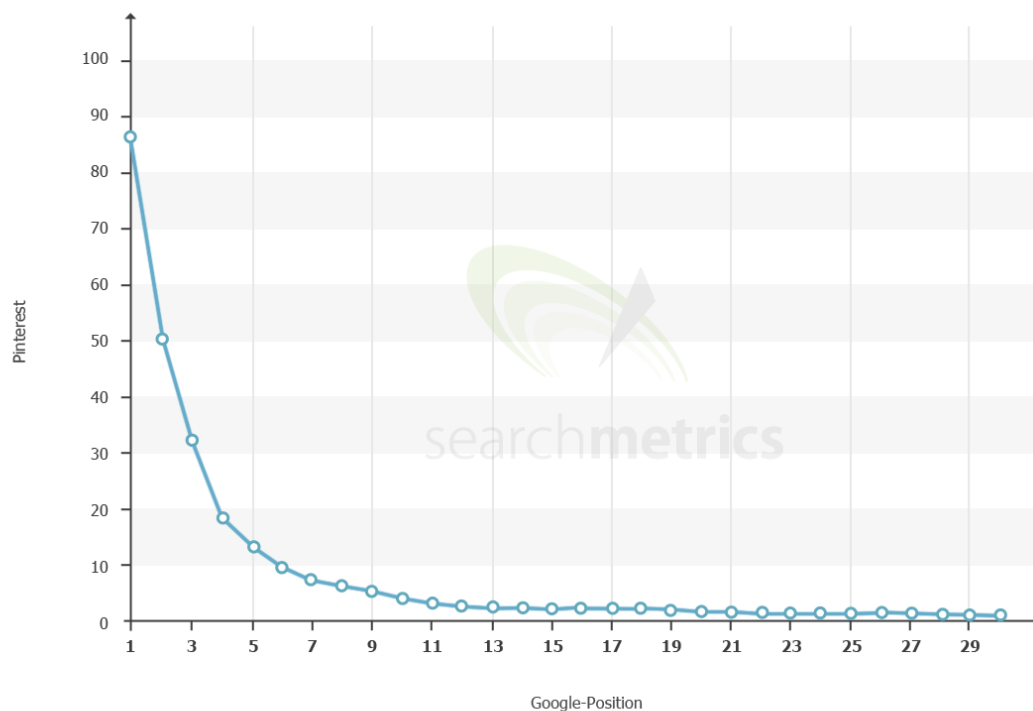


Figure 36: Average by ranking – Pinterest signals

In summary, the following can be said about social media factors:

Typical for these curves is the e-function shape, which is characterized by a sharp drop between the ranks 1 to 2 followed by a much shallower slope as the curve slowly converges to the x-axis. URLs that rank at the top of search results have by far the greatest number of social signals across all the analyzed networks.

4.3 Google+ new in the analysis – with very high correlation

Of particular interest in the analysis of Ranking Factors for the search engine Google is clearly the correlation of signals from the company's own Google+ social network, which has seen an enormous growth in users and activity over the past few months.

Due to insufficient data, it was not possible to make any meaningful statements about this network and its signals as a Ranking Factor last year. In the US, Google+ signals are now just behind tweets in absolute terms.

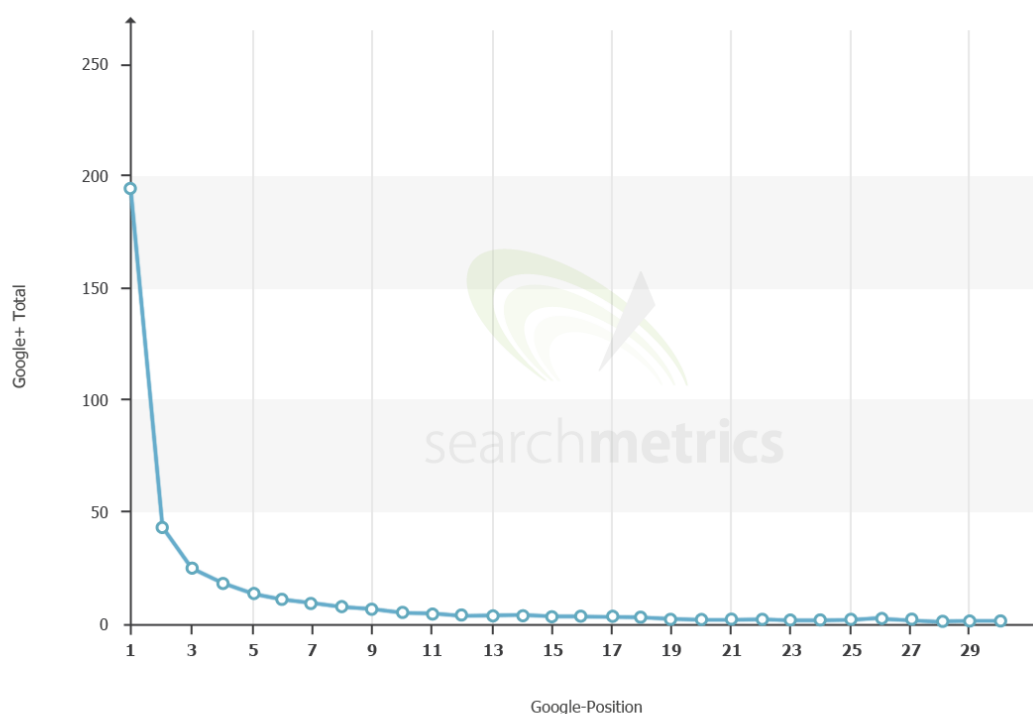


Figure 37: Average by ranking – Google+ signals

Even on a qualitative level, Google+ signals seem to correlate well with top search result rankings. In 2013, plus ones show the highest correlation with good rankings.

4.4 Digression: Google+ will overtake Facebook at the beginning of 2016

At this point we want to offer another digression. This time we will focus on signals and activity of the two by far largest social networks on the Internet.

The catalyst for this analysis is the remarkable development of Google's network Google+, which manifests itself not only in rapid membership growth, but also in the already examined strong correlation with search results, as well as the high number of plus ones for URLs at the top of Google's rankings.

While last year no meaningful conclusions could be drawn with regard to the correlation of plus ones due to lack of data in 2013, it has recorded the highest rank correlation coefficient for this metric.

Searchmetrics has also carried out several detailed studies with respect to the performance and activity of the two networks – Facebook and Google+ – and the interpretation of the respective group strategy behind this accelerated growth –

which will not be presented as part of this study, but will be briefly introduced in the summary of this chapter.

We will, however, briefly introduce one aspect of this study, which focuses on the evolution of networks and their competitive character. This is a forecast, which addresses the following question:

When will Google+ overtake Facebook, the current leader in social media in terms of the number of social signals?

Even the choice of wording of this question hits towards the result – we are not asking "if" but "when."

As a brief introduction, the following can be said about data pools: The starting point for the calculations were the last available monthly percentage growth rates of both networks on average over a period of six months, limited to members' activity in the form of "shares" (Facebook) vs. "plus ones" (Google+). Plotted on a chart, this analysis looks as follows:

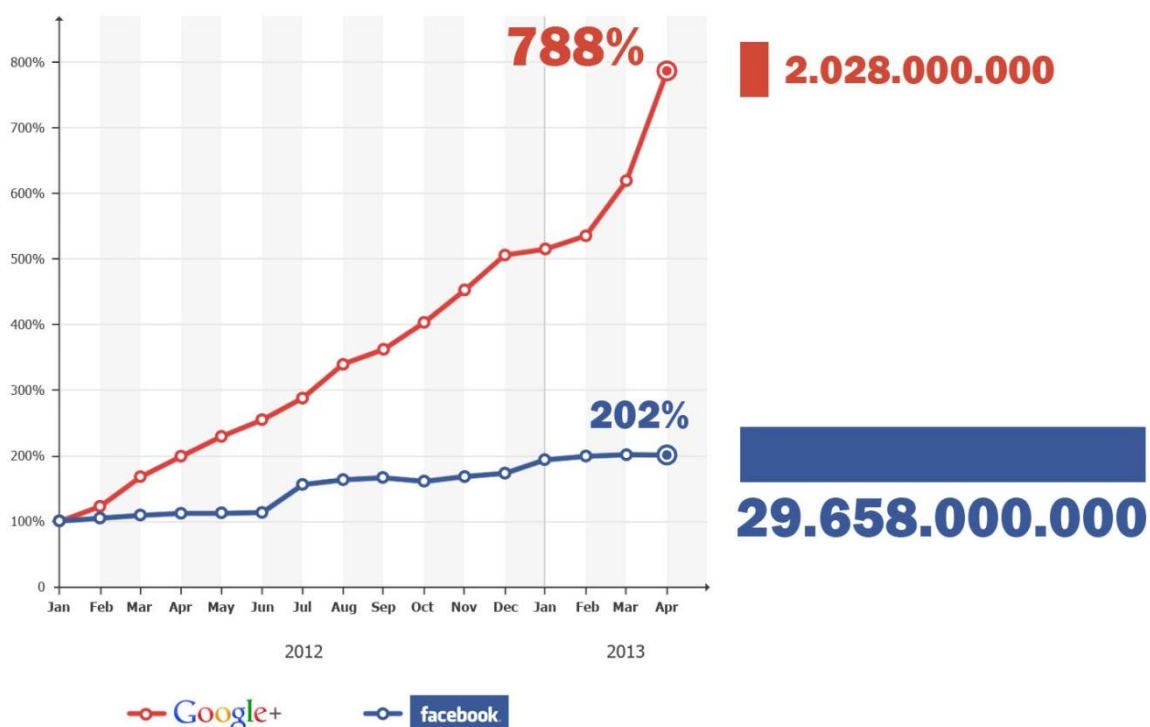


Figure 38: Percentage growth and absolute number of Facebook shares vs plus ones

The growth forecasts for the coming months and years for both networks are based on this data. The purpose is to estimate the point in time when Facebook will be overtaken by Google+ in absolute terms with respect to Facebook shares per month and plus ones, respectively.

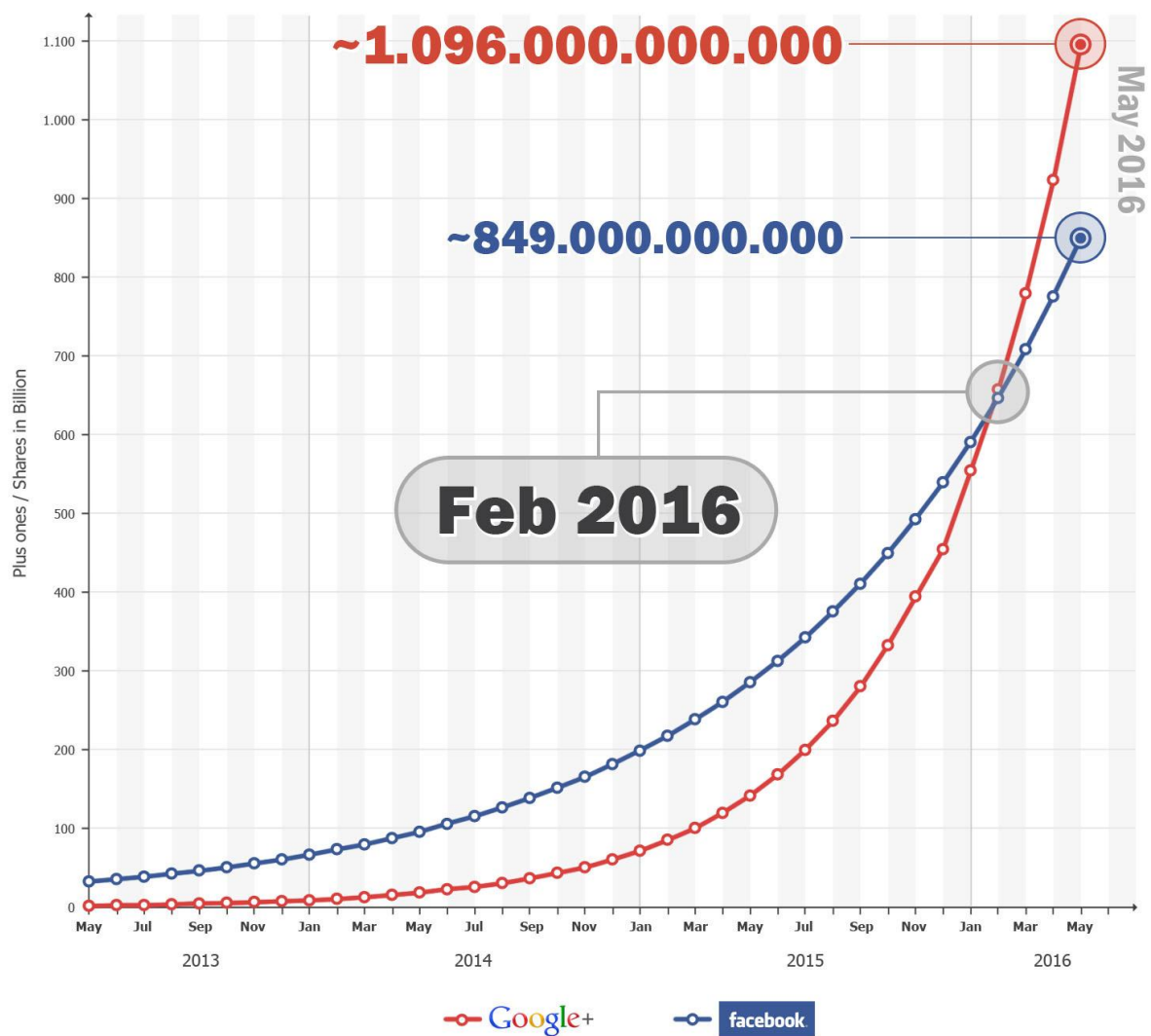


Figure 39: Facebook shares vs. plus ones – growth forecasts in %

Conclusion: According to the projections based on data collected so far, Google + will overtake Facebook in February 2016 based on a comparison of the number of plus ones vs. Facebook shares.

4.5 Summary

Of all the factors, social signals are the strongest in terms of correlation with better rankings. The short to medium-term outlook is clearly positive. These positive results are spreading from Facebook, Google+, Twitter, Pinterest to all networks.

Nevertheless, the question remains as to the extent in which social signals affect rankings directly. As already mentioned, the analyses to date have mainly documented the fact that better ranking URLs have a greater number of social signals than those with a worse ranking.

Several detailed analyses were conducted last year with regard to the causal connection between social signals and good rankings.

For this purpose, two texts¹⁶ on two cognate subjects (four articles in total) that were identical in content were written and published at the same time on the same domain under as neutral conditions as possible (no sitemap, no comments function, no internal / external links from or to the URL). Next, social signals were specifically targeted to the article - under strict network-specific separation. This means, that two articles only received Facebook signals, whereas the copies of this text were exclusively shared on Google+.

While observing the strict separation and compliance with the neutral environment, we evaluated the rankings for specific keywords, which have been determined in advance and were identical for the respective articles.

The results of these social signal studies included: Indexing only through social signals, without presence of any internal or external links, is possible; a significantly smaller number of plus ones is in the position to produce faster indexing or better ranking for an identical set of keywords in a direct comparison; with a repeated increase of shares or plus ones after a period of stagnation, the ranking increases again, even if this has previously also stagnated or even declined.

Without having to reproduce the case studies here in greater detail, we can conclude that it appears that social signals can influence the ranking of URLs even in isolation.

¹⁶ No duplicate content

5. THE BRAND FACTOR

A phenomenon that pervades the entire study, and has had a strong impact on the results in nearly all correlations and detailed charts is the brand factor.

Brands rank in the 1st position, even if they do not meet certain criteria, or if they do not meet them to a sufficient extent. Conclusion: From Google's perspective, brands have a special role and are strongly preferred by being ranked first in SERPs.

5.1 High impact of the brand factor on SERPs

To illustrate the "brand factor" phenomenon we provide again several meaningful example charts, which highlight any distinctive areas in the shape of the curve with respect to the top search result rankings.

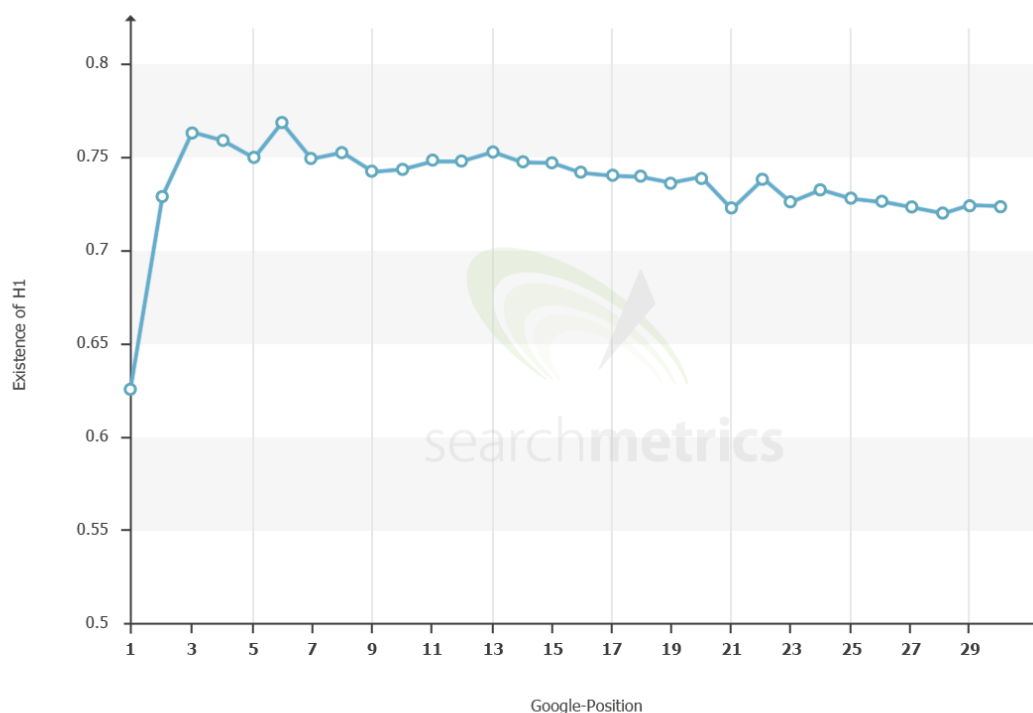


Figure 40: Average by ranking – H1 title tag

The special position of brands is particularly noticeable in on-page coding. It follows from chart 40 that it is much rarer for brand websites to have a H1 title than it is the case for all other websites in the top 30.

Chart 41 shows that it is quite rare for brands to fulfill the factor "keyword in description."

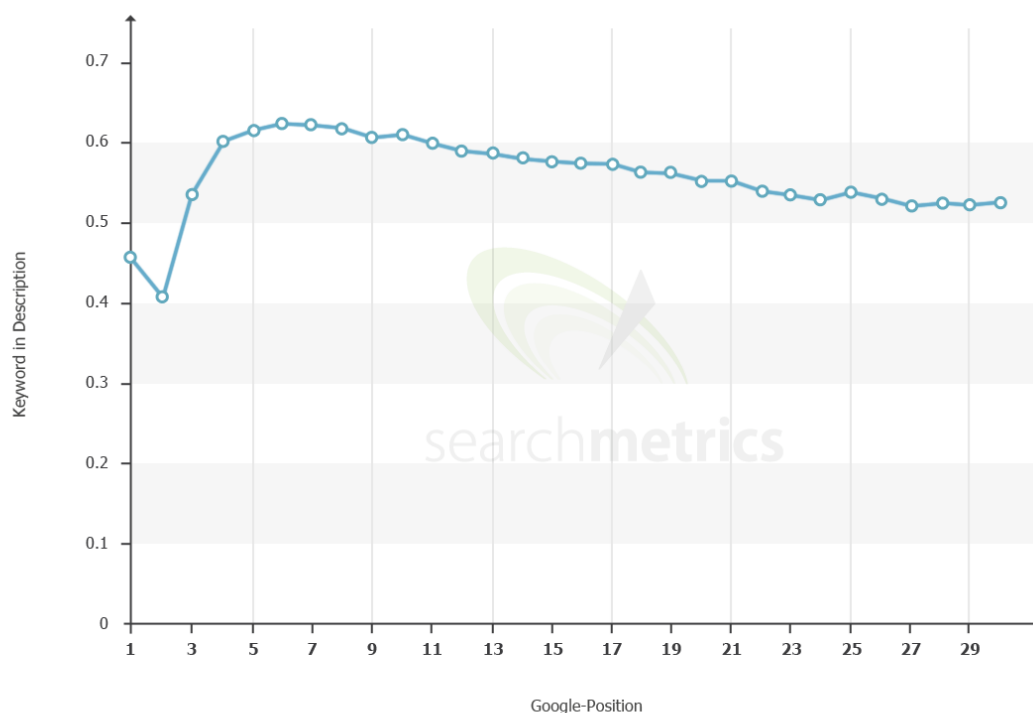


Figure 41: Average by ranking – Keyword in the description

On average, about 55 to 60 percent of webpages out of the top 30 do have a keyword in the description. Only regarding the top 3 ranking pages, the average is significantly lower.

5.2 The top ranks are reserved for brands

While the requirements placed on non-brands to be ranked on the first SERP are consequently, very high, the top search result positions for certain search queries appear to be reserved for brands per se, without the fulfillment or non-fulfillment of certain Ranking Factors being of any relevance.

This failure to implement some important factors, which for non-brands is a basic precondition to be ranked at all, concerns mainly the on-page area – brand websites use i.e. much less frequently keywords in the title or description, offer on average less content and have a lower number of internal links.

On the positive side, brand websites usually feature a better than average backlink profile. Link sources with high SEO Visibility tend to establish links to brands more frequently. In addition, brands receive significantly more social

signals. Consequently, factors from the off-page area seem to have to a considerable impact on the special role of brands.

6. THE SPECIAL ROLE OF WIKIPEDIA

Wikipedia not only has a major impact on ranking of search results but also the correlations calculated for this study. The crowdsourcing project enjoys an extremely high level of trust and high ratings for many keywords (especially "informational keywords"¹⁷).

To generalize, one could say that Wikipedia typically ranks in the 2nd position¹⁸ – for example, behind the respective brand or the specific URL for the search term itself. However, if there is no significant brand that would match the search query, Wikipedia often ranks directly in the 1st position. This is often the case with aforementioned "informational keywords."

6.1 Wikipedia results affect correlations

The actual extent of the Wikipedia influence is highlighted by the following correlations which have been calculated specifically mainly for Wikipedia in addition to the existing correlations.

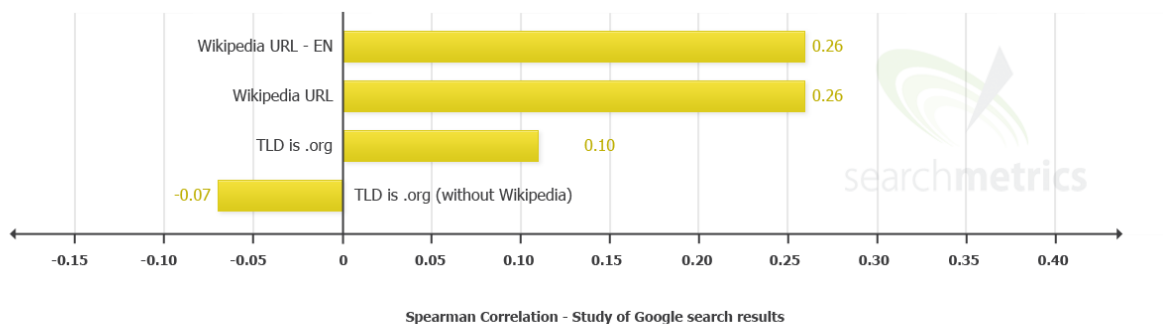


Figure 42: Wikipedia correlation

This chart illustrates the somewhat trivial conclusion that the feature "Wikipedia URL" or "Wikipedia URL-EN" indeed appears to be a Ranking Factor.

¹⁷ Keywords that are googled by users seeking information, e.g. a definition

¹⁸ Taking into account the search volume of a keyword and thus the query frequency

The other two correlations arise from the original intention to analyze the correlation between top-level domains (TLDs) and search result rankings, which, particularly in the case of the "TLD.org" was strongly influenced by other confounding factors and thus no meaningful conclusion could be derived.

At the same time, the individual correlations for "TLD.org", including or excluding Wikipedia, are influenced by a major disruptive factor called Wikipedia, which if added to the mix makes the "TLD.org" factor correlate much better with good rankings than it would be the case without.

The extent of the Wikipedia influence on the search results can be derived from the following detailed correlation for the feature "Wikipedia URL".

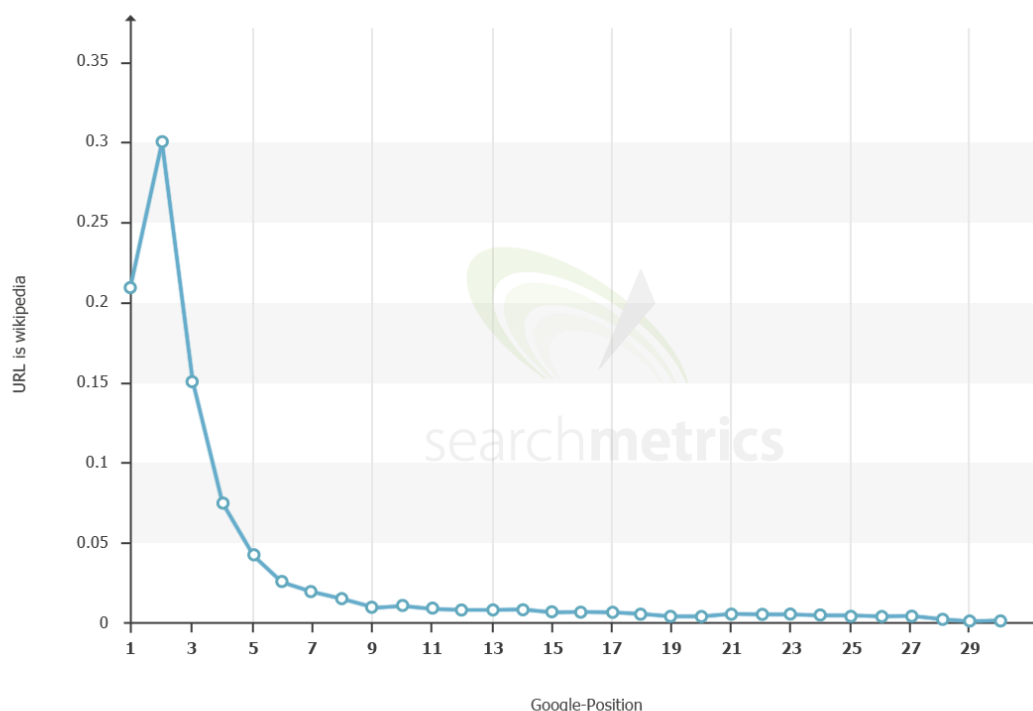


Figure 43: Average ranking – Wikipedia URL

This chart shows that Wikipedia URLs tend to have a high ranking at google.com. Generally, they rank 1st, more often 2nd and rarely 3rd. The data also shows slight clustering at rank 17 and 27.

This unusual data clustering of Wikipedia's results was until recently even more pronounced.

6.2 Rank 17 and 27 "reserved" for Wikipedia results

Not long time ago, the line charts for the individual factors presented in this study showed in many cases substantial breakouts at rank 17 and 27. This is where Wikipedia's results ranked for a long time with very high distribution¹⁹. This regularity was so strongly pronounced that it suggests that it may have been implemented in the algorithm deliberately.

However, after being in place for a long time, this distribution has shifted markedly in the 22nd calendar week due to Penguin 2.0 Update even if not completely dissolved. Nevertheless, the share of keywords to which this distribution applies has decreased significantly in the Searchmetrics database. This would suggest that Google is currently working on its algorithm.

This lead to the fact that the data used in this study was aggregated again in order to take account of the latest developments. The charts in this report are, therefore, based on the current dataset²⁰.

It is noticeable from the 22nd calendar that the Wikipedia results - with a few exceptions - tend to rank in the 1st, 2nd, or 3rd position *or* in the 17th or 27th position. Before the Penguin update, the situation was different. Prior to the 22nd calendar week, there were still combinations of rank 1-3 and 17, 27 for the same keyword in relation to Wikipedia. However, this is now hardly the case.

Meanwhile, Google seems to be moving any Wikipedia results that rank further down to the front and as a result there are often several Wikipedia URLs ranked in the first positions, which are, nevertheless, thematically related to the search query. At the same time, the positions 17 and 27 are often taken by Wikipedia subsidiaries, such as the collection of quotes *Wikiquote*²¹.

The question remains as to what purpose Google has been ranking Wikipedia URLs with such a striking consistency in these parts of SERPs. The fact is that to improve its performance, save processing time and relieve the server, Google caches its SERPs for certain search queries in advance. A possible explanation for the striking ranking distribution would be that Google "parks" Wikipedia results for cache purposes in ranks 17 or 27 to catapult it to the top of search results when e.g. a related, but slightly adapted search query arises. As already mentioned, this clustering has decreased significantly in the meantime.

¹⁹ This has also been the case at rank 37, 47 ff. However, this study focuses on the top 30.

²⁰ June 2013

²¹ <http://www.wikiquote.org/>

Summary of Searchmetrics SEO Ranking Factors – Rank Correlation 2013

To conclude this study, we provide an overview of all rank correlation coefficients as well as their development since the previous year, 2012. The factors which have been newly added in 2013 are grayed out.

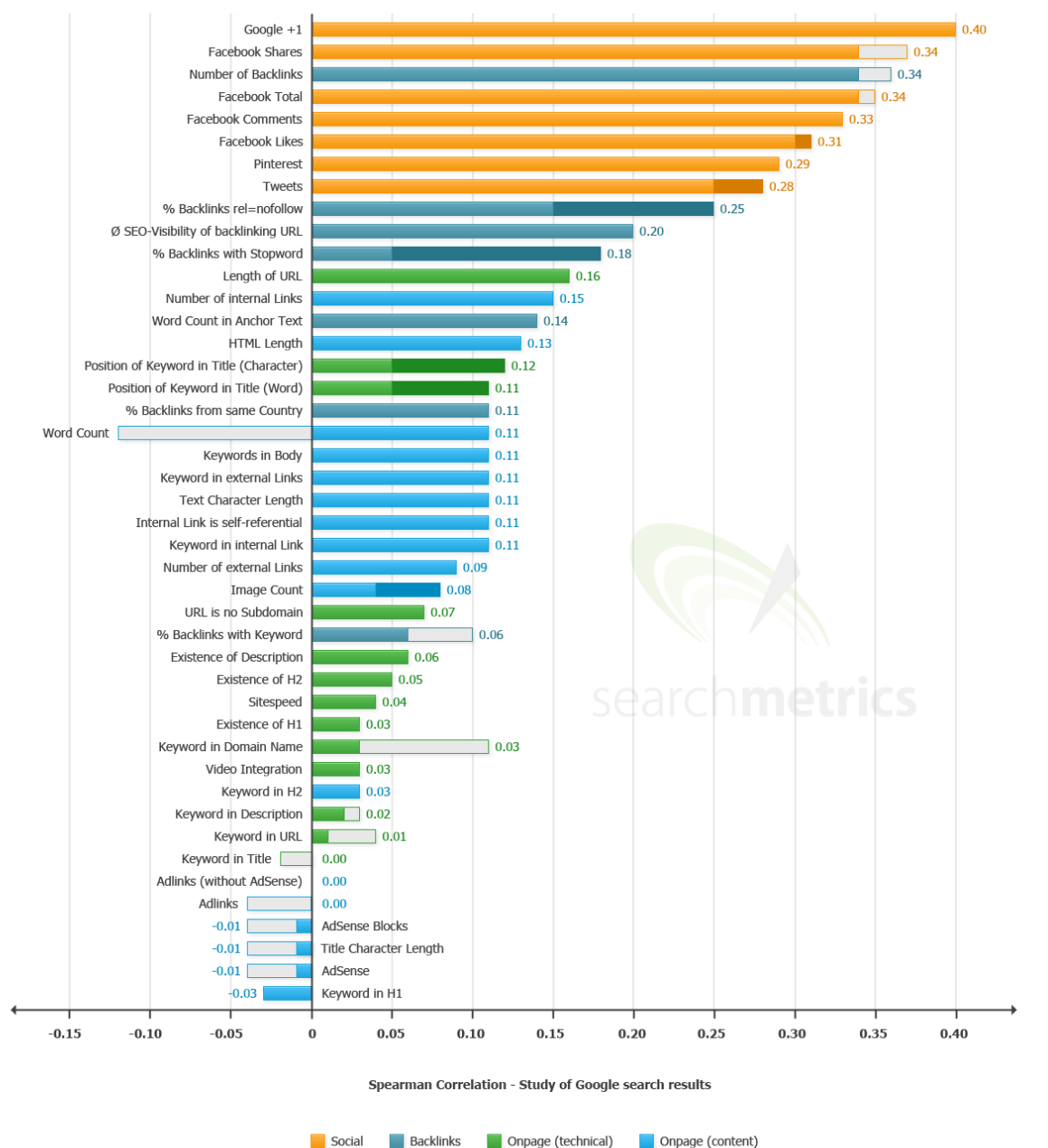


Figure 44: SEO Ranking Factors – Rank Correlation 2013 – Changes since 2012

It is noticeable that social signals correlate well with better rankings and content quality as well as the number and diversity of backlinks seem to have a huge impact on search result rankings.

Summary of most important developments in Ranking Factors – Rank Correlation 2013:

The most important findings of this study are summarized again below:

Keywords are losing in importance – not only with respect to technical on-page factors such as domain names or URLs, but also backlinks in the form of link texts. The days of "hard keyword optimization" of links are well and truly over. In other words, the aforementioned features have lost relevance as a Ranking Factor - Google has simply devalued their once positive influence. However, this does not mean that the basic requirements relating to the existence of keywords and related terms on websites have also become obsolete. It remains one of the basic principles to place relevant keywords on-page, even if not necessarily in the domain or the URL but preferably in the title and description. Google remains a machine, even if the semantics of its algorithm are continuously being optimized. It is also important for a website to contain a description, H1, and H2.

Brands are the exception to the rule. The decoupling and separation of the "brand" factor from the "keyword" factor has been put into practice very well by Google. The majority of top positions in SERPs are occupied by brands and thus websites that mostly do not meet some of the basic principles (often on-page factors).

The entire range of examined social signals correlated well with better rankings. Well-ranked URLs have an abundance of likes, shares, plus ones and tweets. URLs ranked in the 1st position stand out through their large average number of signals. Detailed studies conducted by Searchmetrics suggest that under certain circumstances, social signals appear to have an impact on indexing and ranking of URLs even in isolation. From Google's perspective, social signals function as a reliable signal of good-quality content, especially since the search engine is sitting directly at the source of the social chart in the form of its own network Google+, which is doing very well and is likely to be doing even better in the future.

As already mentioned, several traditional coding factors have gained again in importance. On the content level, nearly all factors correlate significantly better than it was the case in the previous year. It appears that more words in text as well as the integration of additional media in the website both have a positive effect on ranking, while a moderate integration of advertising no longer seems to be an obstacle to good rankings.

In a nutshell: Google favors quality, substance, and ultimately, relevance of the respective search result and always takes into account secondary factors such as brand websites and exceptions, such as Wikipedia.

About Searchmetrics

Searchmetrics is a pioneer and the world's leading provider of search and social analytics software for companies and online agencies.

Searchmetrics offers a unique server infrastructure and software solution which its customers and partners can use to develop intelligent, digital marketing strategies to increase their visibility and success in the digital world. With the world's largest pool of proprietary data and external analyses and information, the Searchmetrics Suite™ uniquely combines various factors, which make it possible to construct accurate forecasts and projections. Furthermore, users can use the software to aggregate and analyze large amounts of data on the ranking of websites, search keywords and the activity of competitors, as well as gain insight into the visibility of websites on search engines in 124 countries, 155 cities and on the largest social media portals. Inbound marketing plays a decisive role in this, as excellent visibility on search engines is a key priority for companies. Only companies that use intelligent and integrated content strategies across all online channels to present themselves will be found and noticed by potential customers, enabling them to increase their ROI. In addition to the Suite, Searchmetrics offers Searchmetrics Essentials, a tool set for getting a quick start in search engine and social network optimization to gain market insights through detailed analysis.

The company has offices in Berlin, London, Paris and New York and sells its SaaS solution through a network of partners and its online store. The fast-growing Berlin-based company, in which the Holtzbrinck publishing group and well-known venture capitalists Neuhaus Partners and Iris Capital hold an interest, currently has over 100 employees and an ever-growing customer base. T-Online, eBay, Siemens, TUI and Kaspersky are some of the well-known customers who already rely on Searchmetrics to achieve excellent search engine rankings and to implement optimal strategies for the digital world.

If you have any questions about the Searchmetrics Suite™ or if you would like to receive a quote:

Contact us now!

For more information, visit our website at: www.searchmetrics.com

Searchmetrics Inc.
1115 Broadway, 12th Floor
New York, NY 10010
Phone: 1 866 411 9494
whitepaper@searchmetrics.com